2.2. Direct methods

By C. Giacovazzo

2.2.1. List of symbols and abbreviations

f_j	atomic scattering factor of jth atom
Z_j	atomic number of jth atom
N	number of atoms in the unit cell
m	order of the point group

$$[\sigma_r]_p, [\sigma_r]_q, [\sigma_r]_N, \ldots = \sum_{j=1}^{p} Z_j^r, \sum_{j=1}^{q} Z_j^r, \sum_{j=1}^{N} Z_j^r, \ldots$$

$[\sigma_r]_N$ is always abbreviated to σ_r when N is the number of atoms in the cell

$$\sum_p, \sum_q, \sum_N, \ldots = \sum_{j=1}^{p} f_j^2, \sum_{j=1}^{q} f_j^2, \sum_{j=1}^{N} f_j^2, \ldots$$

s.f.	structure factor
n.s.f.	normalized structure factor
cs.	centrosymmetric
ncs.	noncentrosymmetric
s.i.	structure invariant
s.s.	structure seminvariant
$\mathbf{C} = (\mathbf{R}, \mathbf{T})$	symmetry operator; $\mathbf{R}$ is the rotational part, $\mathbf{T}$ the translational part
$\varphi_{\mathbf{h}}$	phase of the structure factor $F_{\mathbf{h}} = |F_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}})$

2.2.2. Introduction

Direct methods are today the most widely used tool for solving small crystal structures. They work well both for equal-atom molecules and when a few heavy atoms exist in the structure. In recent years the theoretical background of direct methods has been improved to take into account a large variety of prior information (the form of the molecule, its orientation, a partial structure, the presence of pseudosymmetry or of a superstructure, the availability of isomorphous data or of data affected by anomalous-dispersion effects, ...). Owing to this progress and to the increasing availability of powerful computers, the phase problem for small molecules has been solved in practice: a number of effective, highly automated packages are today available to the scientific community.

The combination of direct methods with so-called direct-space methods have recently allowed the *ab initio* crystal structure solution of proteins. The present limit of complexity is about 2500 non-hydrogen atoms in the asymmetric unit, but diffraction data at atomic resolution (~1 Å) are required. Trials are under way to bring this limit to 1.5 Å and have shown some success.

The theoretical background and tables useful for origin specification are given in Section 2.2.3; in Section 2.2.4 the procedures for normalizing structure factors are summarized. Phase-determining formulae (inequalities, probabilistic formulae for triplet, quartet and quintet invariants, and for one- and two-phase s.s.'s, determinantal formulae) are given in Section 2.2.5. In Section 2.2.6 the connection between direct methods and related techniques in real space is discussed. Practical procedures for solving small-molecule crystal structures are described in Sections 2.2.7 and 2.2.8, and references to the most extensively used packages are given in Section 2.2.9. The integration of direct methods, isomorphous replacement and anomalous-dispersion techniques is briefly discussed in Section 2.2.10.

The reader interested in a more detailed description of the topic is referred to a recent textbook (Giacovazzo, 1998).

2.2.3. Origin specification

(*a*) Once the origin has been chosen, the symmetry operators $\mathbf{C}_s \equiv (\mathbf{R}_s, \mathbf{T}_s)$ and, through them, the algebraic form of the s.f. remain fixed.

A shift of the origin through a vector with coordinates $\mathbf{X}_0$ transforms $\varphi_{\mathbf{h}}$ into

$$\varphi_{\mathbf{h}}' = \varphi_{\mathbf{h}} - 2\pi \mathbf{h} \cdot \mathbf{X}_0 \qquad (2.2.3.1)$$

and the symmetry operators $\mathbf{C}_s$ into $\mathbf{C}_s' = (\mathbf{R}_s', \mathbf{T}_s')$, where

$$\mathbf{R}_s' = \mathbf{R}_s; \quad \mathbf{T}_s' = \mathbf{T}_s + (\mathbf{R}_s - \mathbf{I})\mathbf{X}_0 \quad s = 1, 2, \ldots, m. \qquad (2.2.3.2)$$

(*b*) *Allowed* or *permissible origins* (Hauptman & Karle, 1953, 1959) for a given algebraic form of the s.f. are all those points in direct space which, when taken as origin, maintain the same symmetry operators $\mathbf{C}_s$. The allowed origins will therefore correspond to those points having the same symmetry environment in the sense that they are related to the symmetry elements in the same way. For instance, if $\mathbf{T}_s = 0$ for $s = 1, \ldots, 8$, then the allowed origins in *Pmmm* are the eight inversion centres.

To each functional form of the s.f. a set of permissible origins will correspond.

(*c*) A translation between permissible origins will be called a *permissible* or *allowed translation*. Trivial allowed translations correspond to the lattice periods or to their multiples. A change of origin by an allowed translation does not change the algebraic form of the s.f. Thus, according to (2.2.3.2), all origins allowed by a fixed functional form of the s.f. will be connected by translational vectors $\mathbf{X}_p$ such that

$$(\mathbf{R}_s - \mathbf{I})\mathbf{X}_p = \mathbf{V}, \quad s = 1, 2, \ldots, m, \qquad (2.2.3.3)$$

where $\mathbf{V}$ is a vector with zero or integer components.

In centred space groups, an origin translation corresponding to a centring vector $\mathbf{B}_v$ does not change the functional form of the s.f. Therefore all vectors $\mathbf{B}_v$ represent permissible translations. $\mathbf{X}_p$ will then be an allowed translation (Giacovazzo, 1974) not only when, as imposed by (2.2.3.3), the difference $\mathbf{T}_s' - \mathbf{T}_s$ is equal to one or more lattice units, but also when, for any s, the condition

$$(\mathbf{R}_s - \mathbf{I})\mathbf{X}_p = \mathbf{V} + \alpha\mathbf{B}_v, \quad s = 1, 2, \ldots, m; \quad \alpha = 0, 1 \qquad (2.2.3.4)$$

is satisfied.

We will call any set of cs. or ncs. space groups having the same allowed origin translations a Hauptman–Karle group (H–K group). The 94 ncs. primitive space groups, the 62 primitive cs. groups, the 44 ncs. centred space groups and the 30 cs. centred space groups can be collected into 13, 4, 14 and 5 H–K groups, respectively (Hauptman & Karle, 1953, 1956; Karle & Hauptman, 1961; Lessinger & Wondratschek, 1975). In Tables 2.2.3.1–2.2.3.4 the H–K groups are given together with the allowed origin translations.

(*d*) Let us consider a product of structure factors

$$F_{\mathbf{h}_1}^{A_1} \times F_{\mathbf{h}_2}^{A_2} \times \ldots \times F_{\mathbf{h}_n}^{A_n} = \prod_{j=1}^{n} F_{\mathbf{h}_j}^{A_j}$$

$$= \exp\left(i \sum_{j=1}^{n} A_j \varphi_{\mathbf{h}_j}\right) \prod_{j=1}^{n} |F_{\mathbf{h}_j}|^{A_j}, \qquad (2.2.3.5)$$

A_j being integer numbers.

The factor $\sum_{j=1}^{n} A_j \varphi_{\mathbf{h}_j}$ is the phase of the product (2.2.3.5). A *structure invariant* (s.i.) is a product (2.2.3.5) such that

$$\sum_{j=1}^{n} A_j \mathbf{h}_j = 0. \qquad (2.2.3.6)$$

Since $|F_{\mathbf{h}_j}|$ are usually known from experiment, it is often said that s.i.'s are combinations of phases

$$\sum_{j=1}^{n} A_j \varphi_{\mathbf{h}_j}, \qquad (2.2.3.7)$$

for which (2.2.3.6) holds.

F_0, $F_{\mathbf{h}} F_{-\mathbf{h}}$, $F_{\mathbf{h}} F_{\mathbf{k}} F_{\overline{\mathbf{h+k}}}$, $F_{\mathbf{h}} F_{\mathbf{k}} F_{\mathbf{l}} F_{\overline{\mathbf{h+k+l}}}$, $F_{\mathbf{h}} F_{\mathbf{k}} F_{\mathbf{l}} F_{\mathbf{p}} F_{\overline{\mathbf{h+k+l+p}}}$ are examples of s.i.'s for $n = 1, 2, 3, 4, 5$.

The value of any s.i. does not change with an arbitrary shift of the space-group origin and thus it will depend on the crystal structure only.

(*e*) A *structure seminvariant* (s.s.) is a product of structure factors [or a combination of phases (2.2.3.7)] whose value is unchanged when the origin is moved by an allowed translation.

Let $\mathbf{X}_p$'s be the permissible origin translations of the space group. Then the product (2.2.3.5) [or the sum (2.2.3.7)] is an s.s., if, in accordance with (2.2.3.1),

$$\sum_{j=1}^{n} A_j (\mathbf{h}_j \cdot \mathbf{X}_p) = r, \quad p = 1, 2, \ldots \qquad (2.2.3.8)$$

where r is a positive integer, null or a negative integer.

Conditions (2.2.3.8) can be written in the following more useful form (Hauptman & Karle, 1953):

$$\sum_{j=1}^{n} A_j \mathbf{h}_{s_j} \equiv 0 \quad (\mathrm{mod}\ \omega_s), \qquad (2.2.3.9)$$

where $\mathbf{h}_{s_j}$ is the vector seminvariantly associated with the vector $\mathbf{h}_j$ and ω_s is the seminvariant modulus. In Tables 2.2.3.1–2.2.3.4, the reflection $\mathbf{h}_s$ seminvariantly associated with $\mathbf{h} = (h, k, l)$, the seminvariant modulus ω_s and seminvariant phases are given for every H–K group.

The symbol of any group (*cf.* Giacovazzo, 1974) has the structure $\mathbf{h}_s L \omega_s$, where L stands for the lattice symbol. This symbol is underlined if the space group is cs.

By definition, if the class of permissible origin has been chosen, that is to say, if the algebraic form of the symmetry operators has been fixed, then the value of an s.s. does not depend on the origin but on the crystal structure only.

(*f*) Suppose that we have chosen the symmetry operators $\mathbf{C}_s$ and thus fixed the functional form of the s.f.'s and the set of allowed origins. In order to describe the structure in direct space a unique reference origin must be fixed. Thus the phase-determining process must also require a unique permissible origin congruent to the values assigned to the phases. More specifically, at the beginning of the structure-determining process by direct methods we shall assign as many phases as necessary to define a unique origin among those allowed (and, as we shall see, possibly to fix the enantiomorph). From the theory developed so far it is obvious that arbitrary phases can be assigned to one or more s.f.'s if there is at least one allowed origin which, fixed as the origin of the unit cell, will give those phase values to the chosen reflections. The concept of *linear dependence* will help us to fix the origin.

(*g*) n phases $\varphi_{\mathbf{h}_j}$ are *linearly semidependent* (Hauptman & Karle, 1956) when the n vectors $\mathbf{h}_{s_j}$ seminvariantly associated with the $\mathbf{h}_j$ are linearly dependent modulo ω_s, ω_s being the seminvariant modulus of the space group. In other words, when

$$\sum_{j=1}^{n} A_j \mathbf{h}_{s_j} \equiv 0 \quad (\mathrm{mod}\ \omega_s), \qquad A_q \not\equiv 0 \quad (\mathrm{mod}\ \omega_s) \qquad (2.2.3.10)$$

is satisfied. The second condition means that at least one A_q exists that is not congruent to zero modulo each of the components of ω_s. If (2.2.3.10) is not satisfied for any n-set of integers A_j, the phases $\varphi_{\mathbf{h}_j}$ are *linearly semindependent*. If (2.2.3.10) is valid for $n = 1$ and $A = 1$, then $\mathbf{h}_1$ is said to be linearly semidependent and $\varphi_{\mathbf{h}_1}$ is an s.s. It may be concluded that a seminvariant phase is linearly semidependent, and, *vice versa*, that a phase linearly semidependent is an s.s. In Tables 2.2.3.1–2.2.3.4 the allowed variations (which are those due to the allowed origin translations) for the semindependent phases are given for every H–K group. If $\varphi_{\mathbf{h}_1}$ is linearly semindependent its value can be fixed arbitrarily because at least one origin compatible with the given value exists. Once $\varphi_{\mathbf{h}_1}$ is assigned, the necessary condition to be able to fix a second phase $\varphi_{\mathbf{h}_2}$ is that it should be linearly semindependent of $\varphi_{\mathbf{h}_1}$.

Similarly, the necessary condition to be able arbitrarily to assign a third phase $\varphi_{\mathbf{h}_3}$ is that it should be linearly semindependent from $\varphi_{\mathbf{h}_1}$ and $\varphi_{\mathbf{h}_2}$.

In general, the number of linearly semindependent phases is equal to the dimension of the seminvariant vector ω_s (see Tables 2.2.3.1–2.2.3.4). The reader will easily verify in $(h, k, l) P (2, 2, 2)$ that the three phases $\varphi_{\mathrm{oee}}, \varphi_{\mathrm{eoe}}, \varphi_{\mathrm{eoo}}$ define the origin (o indicates odd, e even).

(*h*) From the theory summarized so far it is clear that a number of semindependent phases $\varphi_{\mathbf{h}_j}$, equal to the dimension of the seminvariant vector ω_s, may be arbitrarily assigned in order to fix the origin. However, it is not always true that only one allowed origin compatible with the given phases exists. An additional condition is required such that only one permissible origin should lie at the intersection of the lattice planes corresponding to the origin-fixing reflections (or on the lattice plane $\mathbf{h}$ if one reflection is sufficient to define the origin). It may be shown that the condition is verified if the determinant formed with the vectors seminvariantly associated with the origin reflections, reduced modulo ω_s, has the value ± 1. In other words, such a determinant should be primitive modulo ω_s.

For example, in $P\bar{1}$ the three reflections

$$\mathbf{h}_1 = (345), \quad \mathbf{h}_2 = (139), \quad \mathbf{h}_3 = (784)$$

define the origin uniquely because

$$\begin{vmatrix} 3 & 4 & 5 \\ 1 & 3 & 9 \\ 7 & 8 & 4 \end{vmatrix} \xrightarrow{\text{reduced mod (2,2,2)}} \begin{vmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{vmatrix} = -1.$$

Furthermore, in $P4mm$ $[\mathbf{h}_s = (h + k, l), \omega_s = (2, 0)]$

$$\mathbf{h}_1 = (5, 2, 0), \quad \mathbf{h}_2 = (6, 2, 1)$$

define the origin uniquely since

$$\begin{vmatrix} 7 & 0 \\ 8 & 1 \end{vmatrix} \xrightarrow{\text{reduced mod (2,0)}} \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

(*i*) If an s.s. or an s.i. has a general value φ for a given structure, it will have a value $-\varphi$ for the enantiomorph structure. If $\varphi = 0, \pi$ the s.s. has the same value for both enantiomorphs. Once the origin has been assigned, in ncs. space groups the sign of a given s.s. $\varphi \neq 0, \pi$ can be assigned to fix the enantiomorph. In practice it is often advisable to use an s.s. or an s.i. whose value is as near as possible to $\pm \pi/2$.

2.2.4. Normalized structure factors

2.2.4.1. Definition of normalized structure factor

The normalized structure factors E (see also Chapter 2.1) are calculated according to (Hauptman & Karle, 1953)

$$|E_{\mathbf{h}}|^2 = |F_{\mathbf{h}}|^2 / \langle |F_{\mathbf{h}}|^2 \rangle, \qquad (2.2.4.1)$$

216

Table 2.2.3.1. *Allowed origin translations, seminvariant moduli and phases for centrosymmetric primitive space groups*

	H–K group					
	$(h, k, l)\underline{P}(2, 2, 2)$		$(h + k, l)\underline{P}(2, 2)$		$(l)\underline{P}(2)$	$(h + k + l)\underline{P}(2)$
Space group	$P\bar{1}$	$Pmna$	$P\dfrac{4}{m}$	$P\dfrac{4}{n}mm$	$P\bar{3}$	$R\bar{3}$
	$P\dfrac{2}{m}$	$Pcca$	$P\dfrac{4_2}{m}$	$P\dfrac{4}{n}cc$	$P\bar{3}1m$	$R\bar{3}m$
	$P\dfrac{2_1}{m}$	$Pbam$	$P\dfrac{4}{n}$	$P\dfrac{4_2}{m}mc$	$P\bar{3}1c$	$R\bar{3}c$
	$P\dfrac{2}{c}$	$Pccn$	$P\dfrac{4_2}{n}$	$P\dfrac{4_2}{m}cm$	$P\bar{3}m1$	$Pm\bar{3}$
	$P\dfrac{2_1}{c}$	$Pbcm$	$P\dfrac{4}{m}mm$	$P\dfrac{4_2}{n}bc$	$P\bar{3}c1$	$Pn\bar{3}$
	$Pmmm$	$Pnnm$	$P\dfrac{4}{m}cc$	$P\dfrac{4_2}{n}nm$	$P\dfrac{6}{m}$	$Pa\bar{3}$
	$Pnnn$	$Pmmn$	$P\dfrac{4}{n}bm$	$P\dfrac{4_2}{m}bc$	$P\dfrac{6_3}{m}$	$Pm\bar{3}m$
	$Pccm$	$Pbcn$	$P\dfrac{4}{n}nc$	$P\dfrac{4_2}{m}nm$	$P\dfrac{6}{m}mm$	$Pn\bar{3}n$
	$Pban$	$Pbca$	$P\dfrac{4}{m}bm$	$P\dfrac{4_2}{n}mc$	$P\dfrac{6}{m}cc$	$Pm\bar{3}n$
	$Pmma$	$Pnma$	$P\dfrac{4}{m}nc$	$P\dfrac{4_2}{n}cm$	$P\dfrac{6_3}{m}cm$	$Pn\bar{3}m$
	$Pnna$				$P\dfrac{6_3}{m}mc$	
Allowed origin translations	$(0, 0, 0);$ $(\frac{1}{2}, 0, 0);$ $(0, \frac{1}{2}, 0);$ $(0, 0, \frac{1}{2});$	$(0, \frac{1}{2}, \frac{1}{2})$ $(\frac{1}{2}, 0, \frac{1}{2})$ $(\frac{1}{2}, \frac{1}{2}, 0)$ $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	$(0, 0, 0)$ $(0, 0, \frac{1}{2})$ $(\frac{1}{2}, \frac{1}{2}, 0)$ $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$		$(0, 0, 0)$ $(0, 0, \frac{1}{2})$	$(0, 0, 0)$ $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$
Vector $\mathbf{h}_s$ seminvariantly associated with $\mathbf{h} = (h, k, l)$	(h, k, l)		$(h + k, l)$		(l)	$(h + k + l)$
Seminvariant modulus $\boldsymbol{\omega}_s$	$(2, 2, 2)$		$(2, 2)$		(2)	(2)
Seminvariant phases	φ_{eee}		$\varphi_{\text{eee}}; \varphi_{\text{ooe}}$		$\varphi_{\text{eee}}; \varphi_{\text{eoe}}$ $\varphi_{\text{oee}}; \varphi_{\text{ooe}}$	$\varphi_{\text{eee}}; \varphi_{\text{ooe}}$ $\varphi_{\text{oeo}}; \varphi_{\text{eoo}}$
Number of semindependent phases to be specified	3		2		1	1

where $|F_{\mathbf{h}}|^2$ is the squared observed structure-factor magnitude on the absolute scale and $\langle |F_{\mathbf{h}}|^2 \rangle$ is the expected value of $|F_{\mathbf{h}}|^2$.

$\langle |F_{\mathbf{h}}|^2 \rangle$ depends on the available *a priori* information. Often, but not always, this may be considered as a combination of several typical situations. We mention:

(*a*) No structural information. The atomic positions are considered random variables. Then

$$\langle |F_{\mathbf{h}}|^2 \rangle = \varepsilon_{\mathbf{h}} \sum_{j=1}^{N} f_j^2 = \varepsilon_{\mathbf{h}} \sum_N$$

so that

$$E_{\mathbf{h}} = \frac{F_{\mathbf{h}}}{(\varepsilon_{\mathbf{h}} \sum_N)^{1/2}}. \qquad (2.2.4.2)$$

$\varepsilon_{\mathbf{h}}$ takes account of the effect of space-group symmetry (see Chapter 2.1).

(*b*) P atomic groups having a known configuration but with unknown orientation and position (Main, 1976). Then a certain number of interatomic distances $r_{j_1 j_2}$ are known and

$$\langle |F_{\mathbf{h}}|^2 \rangle = \varepsilon_{\mathbf{h}} \left(\sum_N + \sum_{i=1}^{P} \sum_{j_1 \neq j_2 = 1}^{M_i} f_{j_1} f_{j_2} \frac{\sin 2\pi q r_{j_1 j_2}}{2\pi q r_{j_1 j_2}} \right),$$

where M_i is the number of atoms in the ith molecular fragment and $q = |\mathbf{h}|$.

(*c*) P atomic groups with a known configuration, correctly oriented, but with unknown position (Main, 1976). Then a certain group of interatomic vectors $\mathbf{r}_{j_1 j_2}$ is fixed and

$$\langle |F_{\mathbf{h}}|^2 \rangle = \varepsilon_{\mathbf{h}} \left(\sum_N + \sum_{i=1}^{P} \sum_{j_1 \neq j_2 = 1}^{M_i} f_{j_1} f_{j_2} \exp 2\pi i \mathbf{h} \cdot \mathbf{r}_{j_1 j_2} \right).$$

The above formula has been derived on the assumption that primitive positional random variables are uniformly distributed over the unit cell. Such an assumption may be considered unfavourable (Giacovazzo, 1988) in space groups for which the allowed shifts of origin, consistent with the chosen algebraic form for the symmetry operators $\mathbf{C}_s$, are arbitrary displacements along any polar axes. Thanks to the indeterminacy in the choice of origin, the first of the shifts $\boldsymbol{\tau}_i$ (to be applied to the ith fragment in order to translate atoms in the correct positions) may be restricted to a region which is smaller than the unit cell (*e.g.* in $P2$ we are free to specify the origin along the diad axis by restricting $\boldsymbol{\tau}_1$ to the family of vectors $\{\boldsymbol{\tau}_1\}$ of type $[x0z]$). The practical consequence is that $\langle |F_{\mathbf{h}}|^2 \rangle$ is significantly modified in polar space groups if $\mathbf{h}$ satisfies

$$\mathbf{h} \cdot \boldsymbol{\tau}_1 = 0,$$

where $\boldsymbol{\tau}_1$ belongs to the family of restricted vectors $\{\boldsymbol{\tau}_1\}$.

(*d*) Atomic groups correctly positioned. Then (Main, 1976; Giacovazzo, 1983*a*)

$$\langle|F_{\mathbf{h}}|^2\rangle = |F_{\mathrm{p},\mathbf{h}}|^2 + \varepsilon_{\mathbf{h}}\sum_q,$$

where $F_{\mathrm{p},\mathbf{h}}$ is the structure factor of the partial known structure and q are the atoms with unknown positions.

(e) A pseudotranslational symmetry is present. Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots$ be the pseudotranslation vectors of order $n_1, n_2, n_3, \ldots$, respectively. Furthermore, let p be the number of atoms (symmetry equivalents included) whose positions are related by pseudotranslational symmetry and q the number of atoms (symmetry equivalents included) whose positions are not related by any pseudotranslation. Then (Cascarano *et al.*, 1985*a*,*b*)

$$\langle|F_{\mathbf{h}}|^2\rangle = \varepsilon_{\mathbf{h}}(\zeta_{\mathbf{h}}\sum_p + \sum_q),$$

where

$$\zeta_{\mathbf{h}} = \frac{(n_1 n_2 n_3 \ldots)\gamma_{\mathbf{h}}}{m}$$

and $\gamma_{\mathbf{h}}$ is the number of times for which algebraic congruences

$$\mathbf{h}\cdot\mathbf{R}_s\mathbf{u}_i \equiv 0 \;(\mathrm{mod}\;1) \quad \text{for } i = 1, 2, 3, \ldots$$

are simultaneously satisfied when s varies from 1 to m. If $\gamma_{\mathbf{h}} = 0$ then $F_{\mathbf{h}}$ is said to be a *superstructure reflection*, otherwise it is a *substructure reflection*.

Often substructures are not ideal: *e.g.* atoms related by pseudotranslational symmetry are ideally located but of different type (replacive deviations from ideality); or they are equal but not ideally located (displacive deviations); or a combination of the two situations occurs. In these cases a correlation exists between the substructure and the superstructure. It has been shown (Mackay, 1953; Cascarano *et al.*, 1988*a*) that the scattering power of the substructural part may be estimated *via* a statistical analysis of diffraction data for ideal pseudotranslational symmetry or for displacive deviations from it, while it is not estimable in the case of replacive deviations.

2.2.4.2. Definition of quasi-normalized structure factor

When probability theory is not used, the *quasi-normalized structure factors* $\mathscr{E}_{\mathbf{h}}$ and the *unitary structure factors* $U_{\mathbf{h}}$ are often used. $\mathscr{E}_{\mathbf{h}}$ and $U_{\mathbf{h}}$ are defined according to

$$|\mathscr{E}_{\mathbf{h}}|^2 = \varepsilon_{\mathbf{h}}|E_{\mathbf{h}}|^2$$

$$U_{\mathbf{h}} = F_{\mathbf{h}}\Big/\left(\sum_{j=1}^N f_j\right).$$

Since $\sum_{j=1}^N f_j$ is the largest possible value for $F_{\mathbf{h}}$, $U_{\mathbf{h}}$ represents the fraction of $F_{\mathbf{h}}$ with respect to its largest possible value. Therefore

Table 2.2.3.2. *Allowed origin translations, seminvariant moduli and phases for noncentrosymmetric primitive space groups*

	H–K group					
	$(h, k, l)P(0, 0, 0)$	$(h, k, l)P(2, 0, 2)$	$(h, k, l)P(0, 2, 0)$	$(h, k, l)P(2, 2, 2)$	$(h, k, l)P(2, 2, 0)$	$(h + k, l)P(2, 0)$
Space group	$P1$	$P2$ $P2_1$	Pm Pc	$P222$ $P222_1$ $P2_12_12$ $P2_12_12_1$	$Pmm2$ $Pmc2_1$ $Pcc2$ $Pma2$ $Pca2_1$ $Pnc2$ $Pmn2_1$ $Pba2$ $Pna2_1$ $Pnn2$	$P4$ $P4_1$ $P4_2$ $P4_3$ $P4mm$ $P4bm$ $P4_2cm$ $P4_2nm$ $P4cc$ $P4nc$ $P4_2mc$ $P4_2bc$
Allowed origin translations	(x, y, z)	$(0, y, 0)$ $(0, y, \frac{1}{2})$ $(\frac{1}{2}, y, 0)$ $(\frac{1}{2}, y, \frac{1}{2})$	$(x, 0, z)$ $(x, \frac{1}{2}, z)$	$(0, 0, 0)$ $(\frac{1}{2}, 0, 0)$ $(0, \frac{1}{2}, 0)$ $(0, 0, \frac{1}{2})$ $(0, \frac{1}{2}, \frac{1}{2})$ $(\frac{1}{2}, 0, \frac{1}{2})$ $(\frac{1}{2}, \frac{1}{2}, 0)$ $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	$(0, 0, z)$ $(0, \frac{1}{2}, z)$ $(\frac{1}{2}, 0, z)$ $(\frac{1}{2}, \frac{1}{2}, z)$	$(0, 0, z)$ $(\frac{1}{2}, \frac{1}{2}, z)$
Vector $\mathbf{h}_s$ seminvariantly associated with $\mathbf{h} = (h, k, l)$	(h, k, l)	(h, k, l)	(h, k, l)	(h, k, l)	(h, k, l)	$(h + k, l)$
Seminvariant modulus ω_s	$(0, 0, 0)$	$(2, 0, 2)$	$(0, 2, 0)$	$(2, 2, 2)$	$(2, 2, 0)$	$(2, 0)$
Seminvariant phases	φ_{000}	φ_{e0e}	φ_{0e0}	φ_{eee}	φ_{ee0}	φ_{ee0} φ_{oo0}
Allowed variations for the semindependent phases	$\|\infty\|$	$\|\infty\|, \|2\|$ if $k = 0$	$\|\infty\|, \|2\|$ if $h = l = 0$	$\|2\|$	$\|\infty\|, \|2\|$ if $l = 0$	$\|\infty\|, \|2\|$ if $l = 0$
Number of semindependent phases to be specified	3	3	3	3	3	2

$$0 \leq |U_{\mathbf{h}}| \leq 1.$$

If atoms are equal, then $U_{\mathbf{h}} = \mathscr{E}_{\mathbf{h}}/N^{1/2}$.

2.2.4.3. The calculation of normalized structure factors

N.s.f.'s cannot be calculated by applying (2.2.4.1) to observed s.f.'s because: (a) the observed magnitudes $I_{\mathbf{h}}$ (already corrected for Lp factor, absorption, …) are on a relative scale; (b) $\langle |F_{\mathbf{h}}|^2 \rangle$ cannot be calculated without having estimated the vibrational motion of the atoms.

This is usually obtained by the well known Wilson plot (Wilson, 1942), according to which observed data are divided into ranges of $s^2 = \sin^2 \theta/\lambda^2$ and averages of intensity $\langle I_{\mathbf{h}} \rangle$ are taken in each shell. Reflection multiplicities and other effects of space-group symmetry on intensities must be taken into account when such averages are calculated. The shells are symmetrically over-lapped in order to reduce statistical fluctuations and are restricted so that the number of reflections in each shell is reasonably large. For each shell

$$K \langle I \rangle = \langle |F|^2 \rangle = \langle |F^{\mathrm{o}}|^2 \rangle \exp(-2Bs^2) \qquad (2.2.4.3)$$

should be obtained, where K is the scale factor needed to place X-ray intensities on the absolute scale, B is the overall thermal parameter and $\langle |F^{\mathrm{o}}|^2 \rangle$ is the expected value of $|F|^2$ in which it is assumed that all the atoms are at rest. $\langle |F^{\mathrm{o}}|^2 \rangle$ depends upon the structural information that is available (see Section 2.2.4.1 for some examples).

Equation (2.2.4.3) may be rewritten as

$$\ln \left\{ \frac{\langle I \rangle}{\langle |F^{\mathrm{o}}|^2 \rangle} \right\} = -\ln K - 2Bs^2,$$

which plotted at various s^2 should be a straight line of which the slope $(2B)$ and intercept $(\ln K)$ on the logarithmic axis can be obtained by applying a linear least-squares procedure.

Very often molecular geometries produce perceptible departures from linearity in the logarithmic Wilson plot. However, the more extensive the available a priori information on the structure is, the closer, on the average, are the Wilson-plot curves to their least-squares straight lines.

Accurate estimates of B and K require good strategies (Rogers & Wilson, 1953) for:

(1) treatment of weak measured data. If weak data are set to zero, there will be bias in the statistics. Methods are, however, available (French & Wilson, 1978) that provide an a posteriori estimate of weak (even negative) intensities by means of Bayesian statistics.

Table 2.2.3.2 (cont.)

$(h+k,l)P(2,2)$	$(h-k,l)P(3,0)$	$(2h+4k+3l)P(6)$	$(l)P(0)$	$(l)P(2)$	$(h+k+l)P(0)$	$(h+k+l)P(2)$
$P\bar{4}$	$P3$	$P312$	$P31m$	$P321$	$R3$	$R32$
$P422$	$P3_1$	$P3_112$	$P31c$	$P3_121$	$R3m$	$P23$
$P42_12$	$P3_2$	$P3_212$	$P6$	$P3_221$	$R3c$	$P2_13$
$P4_122$	$P3m1$	$P6$	$P6_1$	$P622$		$P432$
$P4_12_12$	$P3c1$	$P\bar{6}m2$	$P6_5$	$P6_122$		$P4_232$
$P4_222$		$P\bar{6}c2$	$P6_4$	$P6_522$		$P4_332$
$P4_22_12$			$P6_3$	$P6_222$		$P4_132$
$P4_322$			$P6_2$	$P6_422$		$P\bar{4}3m$
$P4_32_12$			$P6mm$	$P6_322$		$P\bar{4}3n$
$P\bar{4}2m$			$P6cc$	$P\bar{6}2m$		
$P\bar{4}2c$			$P6_3cm$	$P\bar{6}2c$		
$P\bar{4}2_1m$			$P6_3mc$			
$P\bar{4}2_1c$						
$P\bar{4}m2$						
$P\bar{4}c2$						
$P\bar{4}b2$						
$P\bar{4}n2$						
$(0,0,0)$	$(0,0,z)$	$(0,0,0)$	$(0,0,z)$	$(0,0,0)$	(x,x,x)	$(0,0,0)$
$(0,0,\frac{1}{2})$	$(\frac{1}{3},\frac{2}{3},z)$	$(0,0,\frac{1}{2})$		$(0,0,\frac{1}{2})$		$(\frac{1}{2},\frac{1}{2},\frac{1}{2})$
$(\frac{1}{2},\frac{1}{2},0)$	$(\frac{2}{3},\frac{1}{3},z)$	$(\frac{1}{3},\frac{2}{3},0)$				
$(\frac{1}{2},\frac{1}{2},\frac{1}{2})$		$(\frac{1}{3},\frac{2}{3},\frac{1}{2})$				
		$(\frac{2}{3},\frac{1}{3},0)$				
		$(\frac{2}{3},\frac{1}{3},\frac{1}{2})$				
$(h+k,l)$	$(h-k,l)$	$(2h+4k+3l)$	(l)	(l)	$(h+k+l)$	$(h+k+l)$
$(2,2)$	$(3,0)$	(6)	(0)	(2)	(0)	(2)
φ_{eee} φ_{ooe}	φ_{hk0} if $h-k=0$ (mod 3)	φ_{hkl} if $2h+4k+3l=0$ (mod 6)	φ_{hk0}	φ_{hke}	$\varphi_{h,k,\bar{h}+\bar{k}}$	φ_{eee}; φ_{ooe} φ_{oeo}; φ_{ooe}
$\|2\|$	$\|\infty\|$, $\|3\|$ if $l=0$	$\|2\|$ if $h \equiv k$ (mod 3) $\|3\|$ if $l \equiv 0$ (mod 2)	$\|\infty\|$	$\|2\|$	$\|\infty\|$	$\|2\|$
2	2	1	1	1	1	1

Table 2.2.3.3. *Allowed origin translations, seminvariant moduli and phases for centrosymmetric non-primitive space groups*

	H–K group				
	$(h, l)\underline{C}(2, 2)$	$(k, l)\underline{I}(2, 2)$	$(h + k + l)\underline{F}(2)$	$(l)\underline{I}(2)$	$\underline{I}$
Space groups	$C\dfrac{2}{m}$	$Immm$	$Fmmm$	$I\dfrac{4}{m}$	$Im\bar{3}$
	$C\dfrac{2}{c}$	$Ibam$	$Fddd$	$I\dfrac{4_1}{a}$	$Ia\bar{3}$
	$Cmcm$	$Ibca$	$Fm\bar{3}$	$I\dfrac{4}{m}mm$	$Im\bar{3}m$
	$Cmca$	$Imma$	$Fd\bar{3}$	$I\dfrac{4}{m}cm$	$Ia\bar{3}d$
	$Cmmm$		$Fm\bar{3}m$	$I\dfrac{4_1}{a}md$	
	$Cccm$		$Fm\bar{3}c$	$I\dfrac{4_1}{a}cd$	
	$Cmma$		$Fd\bar{3}m$		
	$Ccca$		$Fd\bar{3}c$		
Allowed origin translations	$(0, 0, 0)$	$(0, 0, 0)$	$(0, 0, 0)$	$(0, 0, 0)$	$(0, 0, 0)$
	$(0, 0, \frac{1}{2})$	$(0, 0, \frac{1}{2})$	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	$(0, 0, \frac{1}{2})$	
	$(\frac{1}{2}, 0, 0)$	$(0, \frac{1}{2}, 0)$			
	$(\frac{1}{2}, 0, \frac{1}{2})$	$(\frac{1}{2}, 0, 0)$			
Vector $\mathbf{h}_s$ seminvariantly associated with $\mathbf{h} = (h, k, l)$	(h, l)	(k, l)	$(h + k + l)$	(l)	(h, k, l)
Seminvariant modulus ω_s	$(2, 2)$	$(2, 2)$	(2)	(2)	$(1, 1, 1)$
Seminvariant phases	φ_{eee}	φ_{eee}	φ_{eee}	$\varphi_{\text{eoe}}; \varphi_{\text{eee}}$ $\varphi_{\text{ooe}}; \varphi_{\text{oee}}$	All
Number of semindependent phases to be specified	2	2	1	1	0

(2) treatment of missing weak data (Rogers *et al.*, 1955; Vicković & Viterbo, 1979). All unobserved reflections may assume

$$\mu = |F_{\text{o min}}|^2/3 \text{ for cs. space groups}$$
$$\mu = |F_{\text{o min}}|^2/2 \text{ for ncs. space groups,}$$

where the subscript 'o min' refers to the minimum observed intensity.

Once K and B have been estimated, $E_{\mathbf{h}}$ values can be obtained from experimental data by

$$|E_{\mathbf{h}}|^2 = \frac{KI_{\mathbf{h}}}{\langle |F_{\mathbf{h}}^{\text{o}}|^2 \rangle \exp(-2Bs^2)},$$

Table 2.2.3.4. *Allowed origin translations, seminvariant moduli and phases for noncentrosymmetric non-primitive space groups*

	H–K group					
	$(k, l)C(0, 2)$	$(h, l)C(0, 0)$	$(h, l)C(2, 0)$	$(h, l)C(2, 2)$	$(h, l)A(2, 0)$	$(h, l)I(2, 0)$
Space group	$C2$	Cm Cc	$Cmm2$ $Cmc2_1$ $Ccc2$	$C222$ $C222_1$	$Amm2$ $Abm2$ $Ama2$ $Aba2$	$Imm2$ $Iba2$ $Ima2$
Allowed origin translations	$(0, y, 0)$ $(0, y, \frac{1}{2})$	$(x, 0, z)$	$(0, 0, z)$ $(\frac{1}{2}, 0, z)$	$(0, 0, 0)$ $(0, 0, \frac{1}{2})$ $(\frac{1}{2}, 0, 0)$ $(\frac{1}{2}, 0, \frac{1}{2})$	$(0, 0, z)$ $(\frac{1}{2}, 0, z)$	$(0, 0, z)$ $(\frac{1}{2}, 0, z)$
Vector $\mathbf{h}_s$ seminvariantly associated with $\mathbf{h} = (h, k, l)$	(k, l)	(h, l)	(h, l)	(h, l)	(h, l)	(h, l)
Seminvariant modulus ω_s	$(0, 2)$	$(0, 0)$	$(2, 0)$	$(2, 2)$	$(2, 0)$	$(2, 0)$
Seminvariant phases	φ_{e0e}	φ_{0e0}	φ_{ee0}	φ_{eee}	φ_{ee0}	φ_{ee0}
Allowed variations for the semindependent phases	$\|\infty\|, \|2\|$ if $k = 0$	$\|\infty\|$	$\|\infty\|, \|2\|$ if $l = 0$	$\|2\|$	$\|\infty\|, \|2\|$ if $l = 0$	$\|\infty\|, \|2\|$ if $l = 0$
Number of semindependent phases to be specified	2	2	2	2	2	2

where $\langle |F_{\mathbf{h}}^{o}|^2 \rangle$ is the expected value of $|F_{\mathbf{h}}^{o}|^2$ for the reflection $\mathbf{h}$ on the basis of the available *a priori* information.

2.2.4.4. Probability distributions of normalized structure factors

Under some fairly general assumptions (see Chapter 2.1) probability distribution functions for the variable $|E|$ for cs. and ncs. structures are (see Fig. 2.2.4.1)

$$_{\bar{1}}P(|E|)\,d|E| = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{E^2}{2}\right) d|E| \qquad (2.2.4.4)$$

and

$$_{1}P(|E|)\,d|E| = 2|E| \exp(-|E|^2)\,d|E|, \qquad (2.2.4.5)$$

respectively. Corresponding cumulative functions are (see Fig. 2.2.4.2)

$$_{\bar{1}}N(|E|) = \sqrt{\frac{2}{\pi}} \int_{0}^{|E|} \exp\left(-\frac{t^2}{2}\right) dt = \mathrm{erf}\left(\frac{|E|}{\sqrt{2}}\right),$$

$$_{1}N(|E|) = \int_{0}^{|E|} 2t \exp(-t^2)\,dt = 1 - \exp(-|E|^2).$$

Some moments of the distributions (2.2.4.4) and (2.2.4.5) are listed in Table 2.2.4.1. In the absence of other indications for a given crystal structure, a cs. or an ncs. space group will be preferred according to whether the statistical tests yield values closer to column 2 or to column 3 of Table 2.2.4.1.

For further details about the distribution of intensities see Chapter 2.1.

2.2.5. Phase-determining formulae

From the earliest periods of X-ray structure analysis several authors (Ott, 1927; Banerjee, 1933; Avrami, 1938) have tried to determine atomic positions directly from diffraction intensities.
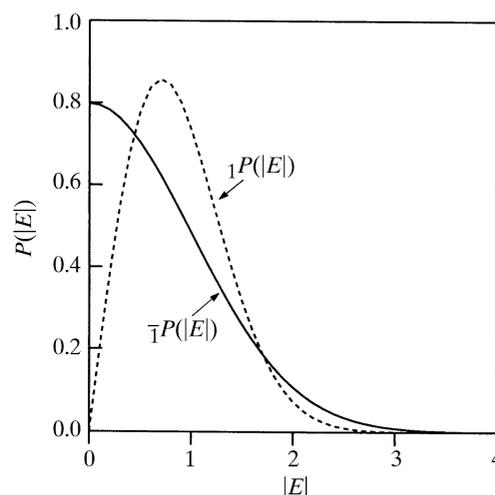
Fig. 2.2.4.1. Probability density functions for cs. and ncs. crystals.

Significant developments are the derivation of inequalities and the introduction of probabilistic techniques *via* the use of joint probability distribution methods (Hauptman & Karle, 1953).

2.2.5.1. Inequalities among structure factors

An extensive system of inequalities exists for the coefficients of a Fourier series which represents a positive function. This can restrict the allowed values for the phases of the s.f.'s in terms of measured structure-factor magnitudes. Harker & Kasper (1948) derived two types of inequalities:

Type 1. A modulus is bound by a combination of structure factors:

$$|U_{\mathbf{h}}|^2 \leq \frac{1}{m} \sum_{s=1}^{m} a_s(-\mathbf{h}) U_{\mathbf{h}(\mathbf{I}-\mathbf{R}_s)}, \qquad (2.2.5.1)$$

where m is the order of the point group and $a_s(-\mathbf{h}) = \exp(-2\pi i \mathbf{h} \cdot \mathbf{T}_s)$.

Table 2.2.3.4 (*cont.*)

$(h, l)I(2, 2)$	$(h+k+l)F(2)$	$(h+k+l)F(4)$	$(l)I(0)$	$(l)I(2)$	$(2k-l)I(4)$	$(l)F(0)$	I
$I222$	$F432$	$F222$	$I4$	$I422$	$I\bar{4}$	$Fmm2$	$I23$
$I2_12_12_1$	$F4_132$	$F23$	$I4_1$	$I4_122$	$I\bar{4}m2$	$Fdd2$	$I2_13$
		$F\bar{4}3m$	$I4mm$	$I\bar{4}2m$	$I\bar{4}c2$		$I432$
		$F\bar{4}3c$	$I4cm$	$I\bar{4}2d$			$I4_132$
			$I4_1md$				$I\bar{4}3m$
			$I4_1cd$				$I\bar{4}3d$
$(0, 0, 0)$	$(0, 0, 0)$	$(0, 0, 0)$	$(0, 0, z)$	$(0, 0, 0)$	$(0, 0, 0)$	$(0, 0, z)$	$(0, 0, 0)$
$(0, 0, \frac{1}{2})$	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$		$(0, 0, \frac{1}{2})$	$(0, 0, \frac{1}{2})$		
$(0, \frac{1}{2}, 0)$		$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$			$(\frac{1}{2}, 0, \frac{3}{4})$		
$(\frac{1}{2}, 0, 0)$		$(\frac{3}{4}, \frac{3}{4}, \frac{3}{4})$			$(\frac{1}{2}, 0, \frac{1}{4})$		
(h, l)	$(h+k+l)$	$(h+k+l)$	(l)	(l)	$(2k-l)$	(l)	(h, k, l)
$(2, 2)$	(2)	(4)	(0)	(2)	(4)	(0)	$(1, 1, 1)$
φ_{eee}	φ_{eee}	φ_{hkl} with $h+k+l \equiv 0$ (mod 4)	φ_{hk0}	φ_{hke}	φ_{hkl} with $(2k-l) \equiv 0$ (mod 4)	φ_{hk0}	All
$\|2\|$	$\|2\|$	$\|2\|$ if $h+k+l \equiv 0$ (mod 2) $\|4\|$ if $h+k+l \equiv 1$ (mod 2)	$\|\infty\|$	$\|2\|$	$\|2\|$ if $h+k+l \equiv 0$ (mod 2) $\|4\|$ if $2k-l \equiv 1$ (mod 2)	$\|\infty\|$	All
2	1	1	1	1	1	1	0

Fig. 2.2.4.2. Cumulative distribution functions for cs. and ncs. crystals.

Table 2.2.4.1. *Moments of the distributions (2.2.4.4) and (2.2.4.5)*

$R(E_s)$ is the percentage of n.s.f.'s with amplitude greater than the threshold E_s.

Criterion	Centrosymmetric distribution	Noncentrosymmetric distribution
$\langle|E|\rangle$	0.798	0.886
$\langle|E|^2\rangle$	1.000	1.000
$\langle|E|^3\rangle$	1.596	1.329
$\langle|E|^4\rangle$	3.000	2.000
$\langle|E|^5\rangle$	6.383	3.323
$\langle|E|^6\rangle$	15.000	6.000
$\langle|E^2-1|\rangle$	0.968	0.736
$\langle(E^2-1)^2\rangle$	2.000	1.000
$\langle(E^2-1)^3\rangle$	8.000	2.000
$\langle|E^2-1|^3\rangle$	8.691	2.415
$R(1)$	0.320	0.368
$R(2)$	0.050	0.018
$R(3)$	0.003	0.0001

Applied to low-order space groups, (2.2.5.1) gives

$$P1: \quad |U_{h,k,l}|^2 \leq 1$$
$$P\bar{1}: \quad U_{h,k,l}^2 \leq 0.5 + 0.5 U_{2h,2k,2l}$$
$$P2_1: \quad |U_{h,k,l}|^2 \leq 0.5 + 0.5(-1)^k U_{2h,0,2l}.$$

The meaning of each inequality is easily understandable: in $P\bar{1}$, for example, $U_{2h,2k,2l}$ must be positive if $|U_{h,k,l}|$ is large enough.

Type 2. The modulus of the sum or of the difference of two structure factors is bound by a combination of structure factors:

$$|U_{\mathbf{h}} \pm U_{\mathbf{h}'}|^2 \leq \frac{1}{m}\left\{ \sum_{s=1}^{m} a_s(-\mathbf{h})U_{\mathbf{h}(\mathbf{I}-\mathbf{R}_s)} + \sum_{s=1}^{m} a_s(-\mathbf{h}')U_{\mathbf{h}'(\mathbf{I}-\mathbf{R}_s)} \right.$$
$$\left. \pm 2\mathscr{Re}\left[\sum_{s=1}^{m} a_s(-\mathbf{h}')U_{\mathbf{h}-\mathbf{h}'\mathbf{R}_s} \right] \right\} \quad (2.2.5.2)$$

where $\mathscr{Re}$ stands for 'real part of'. Equation (2.2.5.2) applied to $P1$ gives

$$|U_{\mathbf{h}} \pm U_{\mathbf{h}'}|^2 \leq 2 \pm 2|U_{\mathbf{h}-\mathbf{h}'}|\cos\varphi_{\mathbf{h}-\mathbf{h}'}.$$

A variant of (2.2.5.2) valid for cs. space groups is

$$(U_{\mathbf{h}} \pm U_{\mathbf{h}'})^2 \leq (1 \pm U_{\mathbf{h}+\mathbf{h}'})(1 \pm U_{\mathbf{h}-\mathbf{h}'}).$$

After Harker & Kasper's contributions, several other inequalities were discovered (Gillis, 1948; Goedkoop, 1950; Okaya & Nitta, 1952; de Wolff & Bouman, 1954; Bouman, 1956; Oda *et al.*, 1961). The most general are the Karle–Hauptman inequalities (Karle & Hauptman, 1950):

$$D_m = \begin{vmatrix} U_0 & U_{-\mathbf{h}_1} & U_{-\mathbf{h}_2} & \cdots & U_{-\mathbf{h}_n} \\ U_{\mathbf{h}_1} & U_0 & U_{\mathbf{h}_1-\mathbf{h}_2} & \cdots & U_{\mathbf{h}_1-\mathbf{h}_n} \\ U_{\mathbf{h}_2} & U_{\mathbf{h}_2-\mathbf{h}_1} & U_0 & \cdots & U_{\mathbf{h}_2-\mathbf{h}_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{\mathbf{h}_n} & U_{\mathbf{h}_n-\mathbf{h}_1} & U_{\mathbf{h}_n-\mathbf{h}_2} & \cdots & U_0 \end{vmatrix} \geq 0. \quad (2.2.5.3)$$

The determinant can be of any order but the leading column (or row) must consist of U's with different indices, although, within the column, symmetry-related U's may occur. For $n = 2$ and $\mathbf{h}_2 = 2\mathbf{h}_1 = 2\mathbf{h}$, equation (2.2.5.3) reduces to

$$D_3 = \begin{vmatrix} U_0 & U_{-\mathbf{h}} & U_{-2\mathbf{h}} \\ U_{\mathbf{h}} & U_0 & U_{-\mathbf{h}} \\ U_{2\mathbf{h}} & U_{\mathbf{h}} & U_0 \end{vmatrix} \geq 0,$$

which, for cs. structures, gives the Harker & Kasper inequality

$$U_{\mathbf{h}}^2 \leq 0.5 + 0.5 U_{2\mathbf{h}}.$$

For $m = 3$, equation (2.2.5.3) becomes

$$D_3 = \begin{vmatrix} U_0 & U_{-\mathbf{h}} & U_{-\mathbf{k}} \\ U_{\mathbf{h}} & U_0 & U_{\mathbf{h}-\mathbf{k}} \\ U_{\mathbf{k}} & U_{\mathbf{k}-\mathbf{h}} & U_0 \end{vmatrix} \geq 0,$$

from which

$$1 - |U_{\mathbf{h}}|^2 - |U_{\mathbf{k}}|^2 - |U_{\mathbf{h}-\mathbf{k}}|^2 + 2|U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}|\cos\alpha_{\mathbf{h},\mathbf{k}} \geq 0, \quad (2.2.5.4)$$

where

$$\alpha_{\mathbf{h},\mathbf{k}} = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}}.$$

If the moduli $|U_{\mathbf{h}}|$, $|U_{\mathbf{k}}|$, $|U_{\mathbf{h}-\mathbf{k}}|$ are large enough, (2.2.5.4) is not satisfied for all values of $\alpha_{\mathbf{h},\mathbf{k}}$. In cs. structures the eventual check that one of the two values of $\alpha_{\mathbf{h},\mathbf{k}}$ does not satisfy (2.2.5.4) brings about the unambiguous identification of the sign of the product $U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}$.

It was observed (Gillis, 1948) that 'there was a number of cases in which both signs satisfied the inequality, one of them by a comfortable margin and the other by only a relatively small margin. In almost all such cases it was the former sign which was the correct one. That suggests that the method may have some power in reserve in the sense that there are still fundamentally stronger inequalities to be discovered'. Today we identify this power in reserve in the use of probability theory.

2.2.5.2. Probabilistic phase relationships for structure invariants

For any space group (see Section 2.2.3) there are linear combinations of phases with cosines that are, in principle, fixed by the $|E|$ magnitudes alone (s.i.'s) or by the $|E|$ values and the trigonometric form of the structure factor (s.s.'s). This result greatly stimulated the calculation of conditional distribution functions

$$P(\Phi|\{R\}), \quad (2.2.5.5)$$

where $R_{\mathbf{h}} = |E_{\mathbf{h}}|$, $\Phi = \sum A_i \varphi_{\mathbf{h}_i}$ is an s.i. or an s.s. and $\{R\}$ is a suitable set of diffraction magnitudes. The method was first proposed by Hauptman & Karle (1953) and was developed further by several authors (Bertaut, 1955*a,b*, 1960; Klug, 1958; Naya *et al.*, 1964, 1965; Giacovazzo, 1980*a*). From a probabilistic point of view the crystallographic problem is clear: the joint distribution $P(E_{\mathbf{h}_1}, \ldots, E_{\mathbf{h}_n})$, from which the conditional distributions (2.2.5.5) can be derived, involves a number of normalized structure factors each of which is a linear sum of random vari-
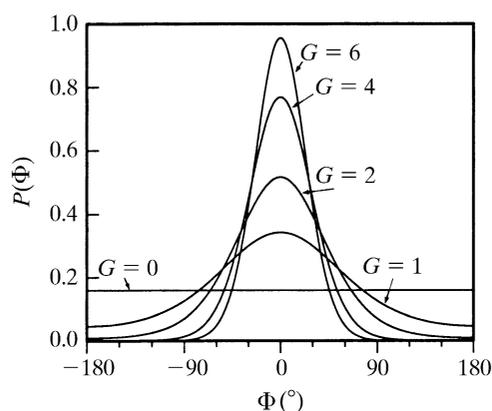
Fig. 2.2.5.1. Curves of (2.2.5.6) for some values of $G = 2\sigma_3\sigma_2^{-3/2}|E_\mathbf{h}E_\mathbf{k}E_{\mathbf{h}-\mathbf{k}}|$.

ables (the atomic contributions to the structure factors). So, for the probabilistic interpretation of the phase problem, the atomic positions and the reciprocal vectors may be considered as random variables. A further problem is that of identifying, for a given Φ, a suitable set of magnitudes $|E|$ on which Φ primarily depends. The formulation of the *nested neighbourhood principle* first (Hauptman, 1975) fixed the idea of defining a sequence of sets of reflections each contained in the succeeding one and having the property that any s.i. or s.s. may be estimated *via* the magnitudes constituting the various neighbourhoods. A subsequent more general theory, the *representation method* (Giacovazzo, 1977a, 1980b), arranges for any Φ the set of intensities in a sequence of subsets in order of their expected effectiveness (in the statistical sense) for the estimation of Φ.

In the following sections the main formulae estimating low-order invariants and seminvariants or relating phases to other phases and diffraction magnitudes are given.

2.2.5.3. Triplet relationships

The basic formula for the estimation of the triplet phase $\Phi = \varphi_\mathbf{h} - \varphi_\mathbf{k} - \varphi_{\mathbf{h}-\mathbf{k}}$ given the parameter $G = 2\sigma_3\sigma_2^{-3/2} \times R_\mathbf{h}R_\mathbf{k}R_{\mathbf{h}-\mathbf{k}}$ is Cochran's (1955) formula

$$P(\Phi) = [2\pi I_0(G)]^{-1}\exp(G\cos\Phi), \qquad (2.2.5.6)$$

where $\sigma_n = \sum_{j=1}^{N} Z_j^n$, Z_j is the atomic number of the jth atom and I_n is the modified Bessel function of order n. In Fig. 2.2.5.1 the distribution $P(\Phi)$ is shown for different values of G.

The conditional probability distribution for $\varphi_\mathbf{h}$, given a set of $(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})$ and $G_j = 2\sigma_3\sigma_2^{-3/2}R_\mathbf{h}R_{\mathbf{k}_j}R_{\mathbf{h}-\mathbf{k}_j}$, is given (Karle & Hauptman, 1956; Karle & Karle, 1966) by

$$P(\varphi_\mathbf{h}) = [2\pi I_0(\alpha)]^{-1}\exp[\alpha\cos(\varphi_\mathbf{h} - \beta_\mathbf{h})], \qquad (2.2.5.7)$$

where

$$\alpha^2 = \left[\sum_{j=1}^{r} G_{\mathbf{h},\mathbf{k}_j}\cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})\right]^2$$
$$+ \left[\sum_{j=1}^{r} G_{\mathbf{h},\mathbf{k}_j}\sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})\right]^2, \qquad (2.2.5.8)$$

$$\tan\beta_\mathbf{h} = \frac{\sum_j G_{\mathbf{h},\mathbf{k}_j}\sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}{\sum_j G_{\mathbf{h},\mathbf{k}_j}\cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}. \qquad (2.2.5.9)$$

$\beta_\mathbf{h}$ is the most probable value for $\varphi_\mathbf{h}$. The variance of $\varphi_\mathbf{h}$ may be obtained from (2.2.5.7) and is given by
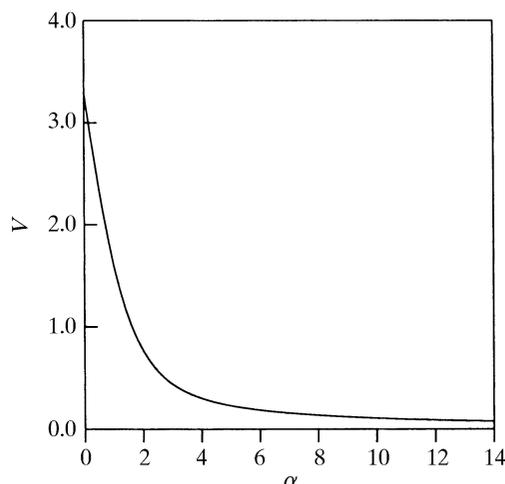

Fig. 2.2.5.2. Variance (in square radians) as a function of α.

$$V_\mathbf{h} = \frac{\pi^2}{3} + [I_0(\alpha)]^{-1}\sum_{n=1}^{\infty}\frac{I_{2n}(\alpha)}{n^2}$$
$$- 4[I_0(\alpha)]^{-1}\sum_{n=0}^{\infty}\frac{I_{2n+1}(\alpha)}{(2n+1)^2}, \qquad (2.2.5.10)$$

which is plotted in Fig. 2.2.5.2.

Equation (2.2.5.9) is the so-called *tangent formula*. According to (2.2.5.10), the larger is α the more reliable is the relation $\varphi_\mathbf{h} = \beta_\mathbf{h}$.

For an equal-atom structure $\sigma_3\sigma_2^{-3/2} = N^{-1/2}$.

The basic conditional formula for sign determination of $E_\mathbf{h}$ in cs. crystals is Cochran & Woolfson's (1955) formula

$$P^+ = \tfrac{1}{2} + \tfrac{1}{2}\tanh\left(\sigma_3\sigma_2^{-3/2}|E_\mathbf{h}|\sum_{j=1}^{r}E_{\mathbf{k}_j}E_{\mathbf{h}-\mathbf{k}_j}\right), \qquad (2.2.5.11)$$

where P^+ is the probability that $E_\mathbf{h}$ is positive and $\mathbf{k}$ ranges over the set of known values $E_\mathbf{k}E_{\mathbf{h}-\mathbf{k}}$. The larger the absolute value of the argument of tanh, the more reliable is the phase indication.

An auxiliary formula exploiting all the $|E|$'s in reciprocal space in order to estimate a single Φ is the $B_{3,0}$ formula (Hauptman & Karle, 1958; Karle & Hauptman, 1958) given by

$$|E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{-\mathbf{h}_1-\mathbf{h}_2}|\cos(\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} - \varphi_{\mathbf{h}_1+\mathbf{h}_2})$$
$$\simeq C\langle(|E_\mathbf{k}|^p - \overline{|E|^p})(|E_{\mathbf{h}_1+\mathbf{k}}|^p - \overline{|E|^p})(|E_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}}|^p - \overline{|E|^p})\rangle_\mathbf{k}$$
$$- \frac{2\sigma_6}{\sigma_4^{3/2}} + \frac{\sigma_8^{1/2}}{\sigma_4}(|E_{\mathbf{h}_1}|^2 + |E_{\mathbf{h}_2}|^2 + |E_{\mathbf{h}_1+\mathbf{h}_2}|^2)\ldots, \qquad (2.2.5.12)$$

where C is a constant which differs for cs. and ncs. crystals, $\overline{|E|^p}$ is the average value of $|E|^p$ and p is normally chosen to be some small number. Several modifications of (2.2.5.12) have been proposed (Hauptman, 1964, 1970; Karle, 1970a; Giacovazzo, 1977b).

A recent formula (Cascarano, Giacovazzo, Camalli *et al.*, 1984) exploits information contained within the second representation of Φ, that is to say, within the collection of special quintets (see Section 2.2.5.6):

$$\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} - \varphi_{\mathbf{h}_1+\mathbf{h}_2} + \varphi_\mathbf{k} - \varphi_\mathbf{k},$$

where $\mathbf{k}$ is a free vector. The formula retains the same algebraic form as (2.2.5.6), but

$$G = \frac{2R_{\mathbf{h}_1}R_{\mathbf{h}_2}R_{\mathbf{h}_3}}{\sqrt{N}}(1 + Q), \qquad (2.2.5.13)$$

where $[\mathbf{h}_3 = -(\mathbf{h}_1 + \mathbf{h}_2)]$,

$$Q = \sum_{\mathbf{k}} \frac{\sum_{i=1}^{'m} A_{\mathbf{k},i}/N}{1 + \left(\varepsilon_{\mathbf{h}_1}\varepsilon_{\mathbf{h}_2}\varepsilon_{\mathbf{h}_3} + \sum_{i=1}^{'m} B_{\mathbf{k},i}\right)\Big/2N},$$

$$A_{\mathbf{k},i} = \varepsilon_{\mathbf{k}}[\varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i}(\varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i})$$
$$+ \varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i}(\varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i})$$
$$+ \varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i}(\varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i})],$$

$$B_{\mathbf{k},i} = \varepsilon_{\mathbf{h}_1}[\varepsilon_{\mathbf{k}}(\varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i})$$
$$+ \varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i}]$$
$$+ \varepsilon_{\mathbf{h}_2}[\varepsilon_{\mathbf{k}}(\varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i})$$
$$+ \varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i}]$$
$$+ \varepsilon_{\mathbf{h}_3}[\varepsilon_{\mathbf{k}}(\varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i})$$
$$+ \varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i}];$$

$\varepsilon = |E|^2 - 1$, $(\varepsilon_{\mathbf{h}_1}\varepsilon_{\mathbf{h}_2}\varepsilon_{\mathbf{h}_3} + \sum_{i=1}^{'m} B_{\mathbf{k},i})$ is assumed to be zero if it is experimentally negative. The prime to the summation warns the reader that precautions have to be taken in order to avoid duplications in the contributions.

G may be positive or negative. In particular, if $G < 0$ the triplet is estimated negative.

The accuracy with which the value of Φ is estimated strongly depends on $\varepsilon_{\mathbf{k}}$. Thus, in practice, only a subset of reciprocal space (the reflections $\mathbf{k}$ with large values of ε) may be used for estimating Φ.

(2.2.5.13) proved to be quite useful in practice. Positive triplet cosines are ranked in order of reliability by (2.2.5.13) markedly better than by Cochran's parameters. Negative estimated triplet cosines may be excluded from the phasing process and may be used as a figure of merit for finding the correct solution in a multisolution procedure.

2.2.5.4. Triplet relationships using structural information

A strength of direct methods is that no knowledge of structure is required for their application. However, when some *a priori* information is available, it should certainly be a weakness of the methods not to make use of this knowledge. The conditional distribution of Φ given $R_{\mathbf{h}}R_{\mathbf{k}}R_{\mathbf{h}-\mathbf{k}}$ and the first three of the five kinds of *a priori* information described in Section 2.2.4.1 is (Main, 1976; Heinermann, 1977*a*)

$$P(\Phi) \simeq \frac{\exp[2QR_1R_2R_3\cos(\Phi - q)]}{2\pi I_0(2QR_1R_2R_3)}, \tag{2.2.5.14}$$

where

$$Q\exp(iq) = \frac{\sum_{i=1}^{p} g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)}{\langle|F_{\mathbf{h}_1}|^2\rangle^{1/2}\langle|F_{\mathbf{h}_2}|^2\rangle^{1/2}\langle|F_{\mathbf{h}_3}|^2\rangle^{1/2}}.$$

$\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ stand for $\mathbf{h}, -\mathbf{k}, -\mathbf{h} + \mathbf{k}$, and R_1, R_2, R_3 for $R_{\mathbf{h}}, R_{\mathbf{k}}, R_{\mathbf{h}-\mathbf{k}}$. The quantities $\langle|F_{\mathbf{h}_i}|^2\rangle$ have been calculated in Section 2.2.4.1 according to different categories: $g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$ is a suitable average of the product of three scattering factors for the *i*th atomic group, p is the number of atomic groups in the cell including those related by symmetry elements. We have the following categories.

(*a*) *No structural information*

(2.2.5.14) then reduces to (2.2.5.6).

(*b*) *Randomly positioned and randomly oriented atomic groups*

Then

$$g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = \sum_{j,k,l} f_j f_k f_l \langle\exp[2\pi i(\mathbf{h}_1 \cdot \mathbf{r}_{kj} + \mathbf{h}_2 \cdot \mathbf{r}_{lj})]\rangle_R,$$

where $\langle\ldots\rangle_R$ means rotational average. The average of the exponential term extends over all orientations of the triangle formed by the atoms *j*, *k* and *l*, and is given (Hauptman, 1965) by

$$B(z, t) = \langle\exp[2\pi i(\mathbf{h} \cdot \mathbf{r} + \mathbf{h}' \cdot \mathbf{r}')]\rangle$$
$$= \left(\frac{\pi}{2z}\right)^{1/2} \sum_{n=0}^{\infty} \frac{t^{2n}}{(n!)^2} J_{(4n+1)/2}(z),$$

where

$$z = 2\pi[q^2r^2 + 2qrq'r'\cos\varphi_q\cos\varphi_r + q'^2r'^2]^{1/2}$$

and

$$t = [2\pi^2 qrq'r'\sin\varphi_q\sin\varphi_r]/z;$$

q, q', r and r' are the magnitudes of $\mathbf{h}$, $\mathbf{h}'$, $\mathbf{r}$ and $\mathbf{r}'$, respectively; φ_q and φ_r are the angles $(\mathbf{h}, \mathbf{h}')$ and $(\mathbf{r}, \mathbf{r}')$, respectively.

(*c*) *Randomly positioned but correctly oriented atomic groups*

Then

$$g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = \sum_{s=1}^{m} \sum_{j,k,l} f_j f_k f_l$$
$$\times \exp[2\pi i(\mathbf{h}_1 \cdot \mathbf{R}_s\mathbf{r}_{kj} + \mathbf{h}_2 \cdot \mathbf{R}_s\mathbf{r}_{lk})],$$

where the summations over *j*, *k*, *l* are taken over all the atoms in the *i*th group.

A modified expression for g_i has to be used in polar space groups for special triplets (Giacovazzo, 1988).

Translation functions [see Chapter 2.3; for an overview, see also Beurskens *et al.* (1987)] are also used to determine the position of a correctly oriented molecular fragment.

Such functions can work in direct space [expressed as Patterson convolutions (Buerger, 1959; Nordman, 1985) or electron-density convolutions (Rossmann *et al.*, 1964; Argos & Rossmann, 1980)] or in reciprocal space [expressed as correlation functions (Crowther & Blow, 1967; Karle, 1972; Langs, 1985) or residual functions (Rae, 1977)]. Both the probabilistic methods and the translation functions are quite efficient tools: the decision as to which one to use is often a personal choice.

(*d*) *Atomic groups correctly positioned*

Let p be the number of atoms with known position, q the number of atoms with unknown position, F_p and F_q the corresponding structure factors.

Tangent recycling methods (Karle, 1970*b*) may be used for recovering the complete crystal structure. The phase $\varphi_{p,\mathbf{h}}$ is accepted in the starting set as a useful approximation of $\varphi_{\mathbf{h}}$ if $|F_{p,\mathbf{h}}| > \eta|F_{\mathbf{h}}|$, where η is the fraction of the total scattering power contained in the fragment and where $|F_{\mathbf{h}}|$ is associated with $|E_{\mathbf{h}}| > 1.5$.

Tangent recycling methods are applied (Beurskens *et al.*, 1979) with greater effectiveness to difference s.f.'s $\Delta F = (|F| - |F_p|)\exp(i\varphi_p)$. The weighted tangent formula uses $\Delta F_{\mathbf{h}}$ values in order to convert them to more probable $F_{q,\mathbf{h}}$ values.

From a probabilistic point of view (Giacovazzo, 1983*a*; Camalli *et al.*, 1985) the distribution of $\varphi_{\mathbf{h}}$, given $E'_{p,\mathbf{h}}$ and some products $(E'_{\mathbf{k}} - E'_{p,\mathbf{k}})(E'_{\mathbf{h}-\mathbf{k}} - E'_{p,\mathbf{h}-\mathbf{k}})$, is the von Mises function

$$P(\varphi_{\mathbf{h}}|\ldots) = [2\pi I_0(\alpha)]^{-1} \exp[\alpha\cos(\varphi_{\mathbf{h}} - \theta_{\mathbf{h}})], \tag{2.2.5.15}$$

where $\theta_{\mathbf{h}}$, the most probable value of $\varphi_{\mathbf{h}}$, is given by

$$\tan\theta_{\mathbf{h}} \simeq \alpha'_2/\alpha'_1, \tag{2.2.5.16}$$
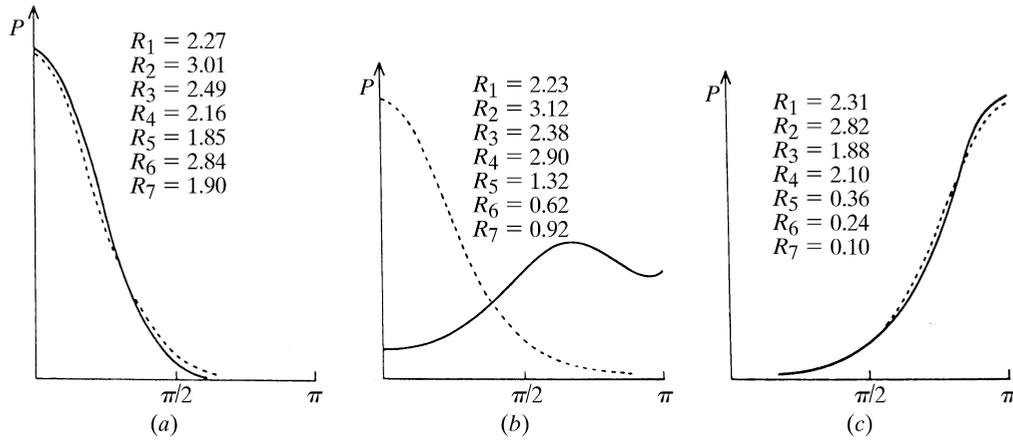$$\alpha^2 = \alpha'^2_1 + \alpha'^2_2$$

and

Fig. 2.2.5.3. Distributions (2.2.5.18) (solid curve) and (2.2.5.20) (dashed curve) for the indicated $|E|$ values in three typical cases.

$$\alpha'_1 = 2R'_{\mathbf{h}}\{\mathscr{R}[E'_{p,\mathbf{h}} + q^{-1/2}\textstyle\sum_{\mathbf{k}}(E'_{\mathbf{k}} - E'_{p,\mathbf{k}})$$
$$\times (E'_{\mathbf{h}-\mathbf{k}} - E'_{p,\mathbf{h}-\mathbf{k}})]\}$$

$$\alpha'_2 = 2R'_{\mathbf{h}}\{\mathscr{I}[E'_{p,\mathbf{h}} + q^{-1/2}\textstyle\sum_{\mathbf{k}}(E'_{\mathbf{k}} - E'_{p,\mathbf{k}})$$
$$\times (E'_{\mathbf{h}-\mathbf{k}} - E'_{p,\mathbf{h}-\mathbf{k}})]\}.$$

$\mathscr{R}$ and $\mathscr{I}$ stand for 'real and imaginary part of', respectively. Furthermore, $E' = F/\sum_q^{1/2}$ is a pseudo-normalized s.f. If no pair $(\varphi_{\mathbf{k}}, \varphi_{\mathbf{h}-\mathbf{k}})$ is known, then

$$\alpha'_1 = 2R'_{\mathbf{h}}R'_{p,\mathbf{h}}\cos\varphi_{p,\mathbf{h}}$$
$$\alpha'_2 = 2R'_{\mathbf{h}}R'_{p,\mathbf{h}}\sin\varphi_{p,\mathbf{h}}$$

and (2.2.5.15) reduces to Sim's (1959) equation

$$P(\varphi_{\mathbf{h}}) \simeq [2\pi I_0(G)]^{-1}\exp[G\cos(\varphi_{\mathbf{h}} - \varphi_{p,\mathbf{h}})], \qquad (2.2.5.17)$$

where $G = 2R'_{\mathbf{h}}R'_{p,\mathbf{h}}$. In this case $\varphi_{p,\mathbf{h}}$ is the most probable value of $\varphi_{\mathbf{h}}$.

(e) *Pseudotranslational symmetry is present*

Substructure and superstructure reflections are then described by different forms of the structure-factor equation (Böhme, 1982; Gramlich, 1984; Fan *et al.*, 1983), so that probabilistic formulae estimating triplet cosines derived on the assumption that atoms are uniformly dispersed in the unit cell cannot hold. In particular, the reliability of each triplet also depends on, besides $R_{\mathbf{h}}, R_{\mathbf{k}}, R_{\mathbf{h}-\mathbf{k}}$, the actual $\mathbf{h}, \mathbf{k}, \mathbf{h} - \mathbf{k}$ indices and on the nature of the pseudotranslation. It has been shown (Cascarano *et al.*, 1985b; Cascarano, Giacovazzo & Luić, 1987) that (2.2.5.7), (2.2.5.8), (2.2.5.9) still hold provided $G_{\mathbf{h},\mathbf{k}_j}$ is replaced by

$$G'_{\mathbf{h},\mathbf{k}_j} = \frac{2R_{\mathbf{h}}R_{\mathbf{k}_j}R_{\mathbf{h}-\mathbf{k}_j}}{\sqrt{N_{\mathbf{h},\mathbf{k}}}},$$

where factors E and n_i are defined according to Section 2.2.4.1,

$$N_{\mathbf{h},\mathbf{k}} = \frac{(\zeta_{\mathbf{h}}[\sigma_2]_p + [\sigma_2]_q)(\zeta_{\mathbf{k}}[\sigma_2]_p + [\sigma_2]_q)(\zeta_{\mathbf{h}-\mathbf{k}}[\sigma_2]_p + [\sigma_2]_q)}{\{(\beta/m)[\sigma_3]_p(n_1^2 n_2^2 n_3^2 \ldots) + [\sigma_3]_q\}^2},$$

and β is the number of times for which

$$\mathbf{h}\mathbf{R}_s \cdot \mathbf{u}_1 \equiv 0 \,(\text{mod } 1) \qquad \mathbf{h}\mathbf{R}_s \cdot \mathbf{u}_2 \equiv 0 \,(\text{mod } 1) \qquad \mathbf{h}\mathbf{R}_s \cdot \mathbf{u}_3 \equiv 0 \,(\text{mod } 1)\ldots$$
$$\mathbf{k}\mathbf{R}_s \cdot \mathbf{u}_1 \equiv 0 \,(\text{mod } 1) \qquad \mathbf{k}\mathbf{R}_s \cdot \mathbf{u}_2 \equiv 0 \,(\text{mod } 1) \qquad \mathbf{k}\mathbf{R}_s \cdot \mathbf{u}_3 \equiv 0 \,(\text{mod } 1)\ldots$$
$$(\mathbf{h} - \mathbf{k})\mathbf{R}_s \cdot \mathbf{u}_1 \equiv 0 \,(\text{mod } 1) \qquad (\mathbf{h} - \mathbf{k})\mathbf{R}_s \cdot \mathbf{u}_2 \equiv 0 \,(\text{mod } 1)$$
$$(\mathbf{h} - \mathbf{k})\mathbf{R}_s \cdot \mathbf{u}_3 \equiv 0 \,(\text{mod } 1)\ldots$$

are simultaneously satisfied when s varies from 1 to m. The above formulae have been generalized (Cascarano *et al.*, 1988b) to the case in which deviations both of replacive and of displacive type from ideal pseudo-translational symmetry occur.

2.2.5.5. Quartet phase relationships

In early papers (Hauptman & Karle, 1953; Simerska, 1956) the phase

$$\Phi = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} - \varphi_{\mathbf{h}+\mathbf{k}+\mathbf{l}}$$

was always expected to be zero. Schenk (1973a,b) [see also Hauptman (1974)] suggested that Φ primarily depends on the seven magnitudes: $R_{\mathbf{h}}, R_{\mathbf{k}}, R_{\mathbf{l}}, R_{\mathbf{h}+\mathbf{k}+\mathbf{l}}$, called *basis magnitudes*, and $R_{\mathbf{h}+\mathbf{k}}, R_{\mathbf{h}+\mathbf{l}}, R_{\mathbf{k}+\mathbf{l}}$, called *cross magnitudes*.

The conditional probability of Φ in $P1$ given seven magnitudes $(R_1 = R_{\mathbf{h}}, \ldots, R_4 = R_{\mathbf{h}+\mathbf{k}+\mathbf{l}}, R_5 = R_{\mathbf{h}+\mathbf{k}}, R_6 = R_{\mathbf{h}+\mathbf{l}}, R_7 = R_{\mathbf{k}+\mathbf{l}})$ according to Hauptman (1975) is

$$P_7(\Phi) = \frac{1}{L}\exp(-2B\cos\Phi)I_0(2\sigma_3\sigma_2^{-3/2}R_5 Y_5)$$
$$\times I_0(2\sigma_3\sigma_2^{-3/2}R_6 Y_6)I_0(2\sigma_3\sigma_2^{-3/2}R_7 Y_7),$$

where L is a suitable normalizing constant which can be derived numerically,

$$B = \sigma_2^{-3}(3\sigma_3^2 - \sigma_2\sigma_4)R_1 R_2 R_3 R_4$$
$$Y_5 = [R_1^2 R_2^2 + R_3^2 R_4^2 + 2R_1 R_2 R_3 R_4\cos\Phi]^{1/2}$$
$$Y_6 = [R_3^2 R_1^2 + R_2^2 R_4^2 + 2R_1 R_2 R_3 R_4\cos\Phi]^{1/2}$$
$$Y_7 = [R_2^2 R_3^2 + R_1^2 R_4^2 + 2R_1 R_2 R_3 R_4\cos\Phi]^{1/2}.$$

For equal atoms $\sigma_2^{-3}(3\sigma_3^2 - \sigma_2\sigma_4) = 2/N$. Denoting

$$C = R_1 R_2 R_3 R_4/N,$$
$$Z_5 = 2Y_5/\sqrt{N}, \qquad Z_6 = 2Y_6/\sqrt{N}, \qquad Z_7 = 2Y_7/\sqrt{N}$$

gives

$$P_7(\Phi) = \frac{1}{L}\exp(-4C\cos\Phi)$$
$$\times I_0(R_5 Z_5)I_0(R_6 Z_6)I_0(R_7 Z_7). \qquad (2.2.5.18)$$

Fig. 2.2.5.3 shows the distribution (2.2.5.18) for three typical cases. It is clear from the figure that the cosine estimated near π or in the middle range will be in poorer agreement with the true values than the cosine near 0 because of the relatively larger values of the variance. In principle, however, the formula is able to estimate negative or enantiomorph-sensitive quartet cosines from the seven magnitudes.

In the cs. case (2.2.5.18) is replaced (Hauptman & Green, 1976) by

Table 2.2.5.1. *List of quartets symmetry equivalent to* $\Phi = \Phi_1$ *in the class mmm*

Quartets	Basis vectors				Cross vectors		
Φ_1	$(1, 2, 3)$	$(\bar{1}, 5, \bar{3})$	$(\bar{1}, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(0, 7, 0)$	$(0, \bar{3}, 11)$	$(\bar{2}, 0, 5)$
Φ_2	$(\bar{1}, 2, 3)$	$(1, 5, \bar{3})$	$(\bar{1}, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(0, 7, 0)$	$(\bar{2}, \bar{3}, 11)$	$(0, 0, 5)$
Φ_3	$(1, 2, \bar{3})$	$(\bar{1}, 5, 3)$	$(\bar{1}, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(0, 7, 0)$	$(0, \bar{3}, 5)$	$(\bar{2}, 0, 11)$
Φ_4	$(\bar{1}, 2, \bar{3})$	$(1, 5, 3)$	$(\bar{1}, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(0, 7, 0)$	$(\bar{2}, \bar{3}, 5)$	$(0, 0, 11)$
Φ_5	$(\bar{1}, 2, 3)$	$(\bar{1}, 5, \bar{3})$	$(1, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(\bar{2}, 7, 0)$	$(0, \bar{3}, 11)$	$(0, 0, 5)$
Φ_6	$(1, 2, 3)$	$(\bar{1}, \bar{5}, \bar{3})$	$(\bar{1}, 5, 8)$	$(1, \bar{2}, \bar{8})$	$(0, \bar{3}, 0)$	$(0, 7, 11)$	$(\bar{2}, 0, 5)$
Φ_7	$(\bar{1}, 2, 3)$	$(1, \bar{5}, \bar{3})$	$(\bar{1}, 5, 8)$	$(1, \bar{2}, \bar{8})$	$(0, \bar{3}, 0)$	$(\bar{2}, 7, 11)$	$(0, 0, 5)$
Φ_8	$(\bar{1}, 2, \bar{3})$	$(\bar{1}, 5, 3)$	$(1, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(\bar{2}, 7, 0)$	$(0, \bar{3}, 5)$	$(0, 0, 11)$
Φ_9	$(1, 2, \bar{3})$	$(\bar{1}, \bar{5}, 3)$	$(\bar{1}, 5, 8)$	$(1, \bar{2}, \bar{8})$	$(0, \bar{3}, 0)$	$(0, 7, 5)$	$(\bar{2}, 0, 11)$
Φ_{10}	$(\bar{1}, 2, \bar{3})$	$(1, \bar{5}, 3)$	$(\bar{1}, 5, 8)$	$(1, \bar{2}, \bar{8})$	$(0, \bar{3}, 0)$	$(\bar{2}, 7, 5)$	$(0, 0, 11)$
Φ_{11}	$(\bar{1}, 2, 3)$	$(\bar{1}, \bar{5}, \bar{3})$	$(1, 5, 8)$	$(1, \bar{2}, \bar{8})$	$(\bar{2}, \bar{3}, 0)$	$(0, 7, 11)$	$(0, 0, 5)$

$$P^{\pm} \simeq \frac{1}{L} \exp(\mp 2C) \cosh(R_5 Z_5^{\pm})$$
$$\times \cosh(R_6 Z_6^{\pm}) \cosh(R_7 Z_7^{\pm}), \qquad (2.2.5.19)$$

where $P^{\pm}$ is the probability that the sign of $E_1 E_2 E_3 E_4$ is positive or negative, and

$$Z_5^{\pm} = \frac{1}{N^{1/2}} (R_1 R_2 \pm R_3 R_4),$$

$$Z_6^{\pm} = \frac{1}{N^{1/2}} (R_1 R_3 \pm R_2 R_4),$$

$$Z_7^{\pm} = \frac{1}{N^{1/2}} (R_1 R_4 \pm R_2 R_3).$$

The normalized probability may be derived by $P^+/(P^+ + P^-)$. More simple probabilistic formulae were derived independently by Giacovazzo (1975, 1976):

$$P_7(\Phi) = [2\pi I_0(G)]^{-1} \exp(G \cos \Phi), \qquad (2.2.5.20)$$

where

$$G = \frac{2C(1 + \varepsilon_5 + \varepsilon_6 + \varepsilon_7)}{1 + Q/(2N)} \qquad (2.2.5.21)$$

$$Q = (\varepsilon_1 \varepsilon_2 + \varepsilon_3 \varepsilon_4)\varepsilon_5 + (\varepsilon_1 \varepsilon_3 + \varepsilon_2 \varepsilon_4)\varepsilon_6 + (\varepsilon_1 \varepsilon_4 + \varepsilon_2 \varepsilon_3)\varepsilon_7$$
$$(2.2.5.22)$$

and $\varepsilon_i = (|E_i|^2 - 1)$. Q is never allowed to be negative.

According to (2.2.5.20) $\cos \Phi$ is expected to be positive or negative according to whether $(\varepsilon_5 + \varepsilon_6 + \varepsilon_7 + 1)$ is positive or negative: the larger is C, the more reliable is the phase indication. For $N \geq 150$, (2.2.5.18) and (2.2.5.20) are practically equivalent in all cases. If N is small, (2.2.5.20) is in good agreement with (2.2.5.18) for quartets strongly defined as positive or negative, but in poor agreement for enantiomorph-sensitive quartets (see Fig. 2.2.5.3).

In cs. cases the sign probability for $E_1 E_2 E_3 E_4$ is

$$P^+ = \tfrac{1}{2} + \tfrac{1}{2}\tanh(G/2), \qquad (2.2.5.23)$$

where G is defined by (2.2.5.21).

All three cross magnitudes are not always in the set of measured reflections. From marginal distributions the following formulae arise (Giacovazzo, 1977c; Heinermann, 1977b):

(a) in the ncs. case, if R_7, or R_6 and R_7, or R_5 and R_6 and R_7, are not in the measurements, then (2.2.5.18) is replaced by

$$P(\Phi | R_1, \dots, R_6) \simeq \frac{1}{L'} \exp(-2C \cos \Phi) I_0(R_5 Z_5) I_0(R_6 Z_6),$$

or

$$P(\Phi | R_1, \dots, R_5) \simeq \frac{1}{L''} I_0(R_5 Z_5),$$

or

$$P(\Phi | R_1, \dots, R_4) \simeq \frac{1}{L'''} \exp(2C \cos \Phi),$$

respectively.

(b) in the same situations, we have for cs. cases

$$P^{\pm} \simeq \frac{1}{L'} \exp(\mp C) \cosh(R_5 Z_5^{\pm}) \cosh(R_6 Z_6^{\pm}),$$

or

$$P^{\pm} \simeq \frac{1}{L''} \cosh(R_5 Z_5^{\pm}),$$

or

$$P^{\pm} = \frac{1}{L'''} \exp(\pm C) \simeq 0.5 + 0.5 \tanh(\pm C),$$

respectively.

Equations (2.2.5.20) and (2.2.5.23) are easily modifiable when some cross magnitudes are not in the measurements. If R_i is not measured then (2.2.5.20) or (2.2.5.23) are still valid provided that in G it is assumed that $\varepsilon_i = 0$. For example, if R_7 and R_6 are not in the data then (2.2.5.21) and (2.2.5.22) become

$$G = \frac{2C(1 + \varepsilon_5)}{1 + Q/(2N)}, \qquad Q = (\varepsilon_1 \varepsilon_2 + \varepsilon_3 \varepsilon_4)\varepsilon_5.$$

In space groups with symmetry higher than $P\bar{1}$ more symmetry-equivalent quartets can exist of the type

$$\psi = \varphi_{\mathbf{h}R_\alpha} + \varphi_{\mathbf{k}R_\beta} + \varphi_{\mathbf{l}R_\gamma} + \varphi_{\overline{(\mathbf{h+k+l})}R_\delta},$$

where $\mathbf{R}_\alpha, \mathbf{R}_\beta, \mathbf{R}_\gamma, \mathbf{R}_\delta$ are rotation matrices of the space group. The set $\{\psi\}$ is called the *first representation* of Φ. In this case Φ primarily depends on more than seven magnitudes. For example, let us consider in $Pmmm$ the quartet

$$\Phi = \varphi_{123} + \varphi_{\bar{1}5\bar{3}} + \varphi_{\bar{1}58} + \varphi_{1\bar{2}\bar{8}}.$$

Quartets symmetry equivalent to Φ and respective cross terms are given in Table 2.2.5.1.

Experimental tests on the application of the representation concept to quartets have been made (Busetta *et al.*, 1980). It was shown that quartets with more than three cross magnitudes are more accurately estimated than other quartets. Also, quartets with a cross reflection which is systematically absent were shown to be of significant importance in direct methods. In this context it is noted that systematically absent reflections are not usually included in the set of diffraction data. This custom, not exceptionable when only triplet relations are used, can give rise to a loss of information when quartets are used. In fact the usual programs of direct methods discard quartets as soon as one of the cross reflections is not measured, so that systematic absences are

dealt with in the same manner as those reflections which are outside the sphere of measurements.

2.2.5.6. Quintet phase relationships

A quintet phase

$$\Phi = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{m}} + \varphi_{\overline{\mathbf{h+k+l+m}}}$$

may be considered as the sum of three suitable triplets or the sum of a triplet and a quartet, *i.e.*

$$\Phi = (\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h+k}}) + (\varphi_{\mathbf{l}} + \varphi_{\mathbf{m}} - \varphi_{\mathbf{l+m}})$$
$$+ (\varphi_{\mathbf{h+k}} + \varphi_{\mathbf{l+m}} + \varphi_{\overline{\mathbf{h+k+l+m}}})$$

or

$$\Phi = (\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h+k}}) + (\varphi_{\mathbf{l}} + \varphi_{\mathbf{m}} + \varphi_{\overline{\mathbf{h+k+l+m}}} + \varphi_{\mathbf{h+k}}).$$

It depends primarily on 15 magnitudes: the five *basis magnitudes*

$$R_{\mathbf{h}}, \quad R_{\mathbf{k}}, \quad R_{\mathbf{l}}, \quad R_{\mathbf{m}}, \quad R_{\mathbf{h+k+l+m}},$$

and the ten *cross magnitudes*

$$R_{\mathbf{h+k}}, \quad R_{\mathbf{h+l}}, \quad R_{\mathbf{h+m}}, \quad R_{\mathbf{k+l+m}}, \quad R_{\mathbf{k+l}},$$
$$R_{\mathbf{k+m}}, \quad R_{\mathbf{h+l+m}}, \quad R_{\mathbf{l+m}}, \quad R_{\mathbf{h+k+m}}, \quad R_{\mathbf{h+k+l}}.$$

In the following we will denote

$$R_1 = R_{\mathbf{h}}, \quad R_2 = R_{\mathbf{k}}, \ldots, \quad R_{15} = R_{\mathbf{h+k+l}}.$$

Conditional distributions of Φ in $P1$ and $P\bar{1}$ given the 15 magnitudes have been derived by several authors and allow in favourable circumstances in ncs. space groups the quintets having Φ near 0 or near π or near $\pm\pi/2$ to be identified. Among others, we remember:

(*a*) the semi-empirical expression for $P_{15}(\Phi)$ suggested by Van der Putten & Schenk (1977):

$$P(\Phi|\ldots) \simeq \frac{1}{L}\exp\left[\left(6 - \sum_{j=6}^{15} R_j^2\right)2C\cos\Phi\right]\prod_{j=6}^{15} I_0(2R_j Y_j),$$

where

$$C = N^{-3/2}R_1 R_2 R_3 R_4 R_5$$

and Y_j is an expression related to the jth of the ten quartets connected with the quintet Φ;

(*b*) the formula by Fortier & Hauptman (1977), valid in $P\bar{1}$, which is able to predict the sign of a quintet by means of an expression which involves a summation over 1024 sets of signs;

(*c*) the expression by Giacovazzo (1977*d*), according to which

$$P_{15}(\Phi) \simeq [2\pi I_0(G)]^{-1}\exp(G\cos\Phi), \qquad (2.2.5.24)$$

where

$$G = \frac{2C}{1 + 6(N)^{1/2}}\left[\frac{1 + A + B}{1 + D/(2N)}\right] \qquad (2.2.5.25)$$

and where

$$A = \sum_{i=6}^{15}\varepsilon_i,$$

$$\begin{aligned}B = {}& \varepsilon_6\varepsilon_{13} + \varepsilon_6\varepsilon_{15} + \varepsilon_6\varepsilon_{14} + \varepsilon_7\varepsilon_{11} + \varepsilon_7\varepsilon_{15} + \varepsilon_7\varepsilon_{12} \\ & + \varepsilon_8\varepsilon_{10} + \varepsilon_8\varepsilon_{14} + \varepsilon_8\varepsilon_{12} + \varepsilon_{10}\varepsilon_{15} + \varepsilon_{10}\varepsilon_9 + \varepsilon_{11}\varepsilon_{14} \\ & + \varepsilon_{11}\varepsilon_9 + \varepsilon_{13}\varepsilon_9 + \varepsilon_{13}\varepsilon_{12},\end{aligned}$$

$$\begin{aligned}D = {}& \varepsilon_1\varepsilon_2\varepsilon_6 + \varepsilon_1\varepsilon_3\varepsilon_7 + \varepsilon_1\varepsilon_4\varepsilon_8 + \varepsilon_1\varepsilon_5\varepsilon_9 + \varepsilon_1\varepsilon_{10}\varepsilon_{15} \\ & + \varepsilon_1\varepsilon_{11}\varepsilon_{14} + \varepsilon_1\varepsilon_{13}\varepsilon_{12} + \varepsilon_2\varepsilon_3\varepsilon_{10} + \varepsilon_2\varepsilon_4\varepsilon_{11} \\ & + \varepsilon_2\varepsilon_5\varepsilon_{12} + \varepsilon_2\varepsilon_7\varepsilon_{15} + \varepsilon_2\varepsilon_8\varepsilon_{14} + \varepsilon_2\varepsilon_{13}\varepsilon_9 + \varepsilon_3\varepsilon_4\varepsilon_{13} \\ & + \varepsilon_3\varepsilon_5\varepsilon_{14} + \varepsilon_3\varepsilon_6\varepsilon_{15} + \varepsilon_3\varepsilon_8\varepsilon_{12} + \varepsilon_3\varepsilon_{11}\varepsilon_9 + \varepsilon_4\varepsilon_5\varepsilon_{15} \\ & + \varepsilon_4\varepsilon_6\varepsilon_{14} + \varepsilon_4\varepsilon_7\varepsilon_{12} + \varepsilon_4\varepsilon_{10}\varepsilon_9 + \varepsilon_5\varepsilon_6\varepsilon_{13} + \varepsilon_5\varepsilon_7\varepsilon_{11} \\ & + \varepsilon_5\varepsilon_8\varepsilon_{10}.\end{aligned}$$

For cs. cases (2.2.5.24) reduces to

$$P^+ \simeq 0.5 + 0.5\tanh(G/2). \qquad (2.2.5.26)$$

Positive or negative quintets may be identified according to whether G is larger or smaller than zero.

If R_i is not measured then (2.2.5.24) and (2.2.5.25) are still valid provided that in (2.2.5.25) $\varepsilon_i = 0$.

If the symmetry is higher than in $P\bar{1}$ then more symmetry-equivalent quintets can exist of the type

$$\psi = \varphi_{\mathbf{h}R_\alpha} + \varphi_{\mathbf{k}R_\beta} + \varphi_{\mathbf{l}R_\gamma} + \varphi_{\mathbf{m}R_\delta} + \varphi_{\overline{(\mathbf{h+k+l+m})}R_\varepsilon},$$

where $\mathbf{R}_\alpha, \ldots, \mathbf{R}_\varepsilon$ are rotation matrices of the space groups. The set $\{\psi\}$ is called the first representation of Φ. In this case Φ primarily depends on more than 15 magnitudes which all have to be taken into account for a careful estimation of Φ (Giacovazzo, 1980*a*).

A wide use of quintet invariants in direct-methods procedures is prevented for two reasons: (*a*) the large correlation of positive quintet cosines with positive triplets; (*b*) the large computing time necessary for their estimation [quintets are phase relationships of order $1/(N\sqrt{N})$, so a large number of quintets have to be estimated in order to pick up a sufficient percentage of reliable ones].

2.2.5.7. Determinantal formulae

In a crystal structure with N identical atoms the joint probability distribution of n normalized s.f.'s $E_{\mathbf{h}_1+\mathbf{k}}, E_{\mathbf{h}_2+\mathbf{k}}, \ldots, E_{\mathbf{h}_n+\mathbf{k}}$ under the following conditions:

(*a*) the structure is kept fixed whereas $\mathbf{k}$ is the primitive random variable;

(*b*) $E_{\mathbf{h}_i-\mathbf{h}_j}$, $i, j = 1, \ldots, n$, have values which are known *a priori*;

is given (Tsoucaris, 1970) [see also Castellano *et al.* (1973) and Heinermann *et al.* (1979)] by

$$P(E_1, E_2, \ldots, E_n) = (2\pi)^{-n/2}D_n^{-1/2}\exp(-\tfrac{1}{2}Q_n) \qquad (2.2.5.27)$$

for cs. structures and

$$P(E_1, E_2, \ldots, E_n) = (2\pi)^{-n}D_n^{-1/2}\exp(-Q_n) \qquad (2.2.5.28)$$

for ncs. structures. In (2.2.5.27) and (2.2.5.28) we have denoted

$$D_n = \lambda, \qquad Q_n = \sum_{p,q=1}^{n}\Lambda_{pq}E_p E_q^*$$

$$E_j = E_{\mathbf{h}_j+\mathbf{k}}, \qquad U_{pq} = U_{\mathbf{h}_p-\mathbf{h}_q}, \qquad j, p, q = 1, \ldots, n.$$

Λ_{pq} is an element of $\boldsymbol{\lambda}^{-1}$, and $\boldsymbol{\lambda}$ is the covariance matrix with elements

$$\langle E_{\mathbf{h}_p+\mathbf{k}} E_{\mathbf{h}_q+\mathbf{k}} \rangle = U_{\mathbf{h}_p-\mathbf{h}_q}$$

$$\lambda = \begin{vmatrix} 1 & U_{12} & \dots & U_{1q} & \dots & U_{1n} \\ U_{21} & 1 & \dots & U_{2q} & \dots & U_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ U_{p1} & U_{p2} & \dots & U_{pq} & \dots & U_{pn} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ U_{n1} & U_{n2} & \dots & U_{nq} & \dots & 1 \end{vmatrix}.$$

λ is a K–H determinant: therefore $D_n \geq 0$. Let us call

$$\Delta_{n+1} = \frac{1}{N} \begin{vmatrix} 1 & U_{12} & \dots & U_{1n} & E_{\mathbf{h}_1+\mathbf{k}} \\ U_{21} & 1 & \dots & U_{2n} & E_{\mathbf{h}_2+\mathbf{k}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ U_{n1} & U_{n2} & \dots & 1 & E_{\mathbf{h}_n+\mathbf{k}} \\ E_{-\mathbf{h}_1-\mathbf{k}} & E_{-\mathbf{h}_2-\mathbf{k}} & \dots & E_{-\mathbf{h}_n-\mathbf{k}} & N \end{vmatrix};$$

the K–H determinant obtained by adding to λ the last column and line formed by $E_1, E_2, \dots, E_n$, and $E_1^*, E_2^*, \dots, E_n^*$, respectively. Then (2.2.5.27) and (2.2.5.28) may be written

$$P(E_1, E_2, \dots, E_n)$$
$$= (2\pi)^{-n/2} D_n^{-1/2} \exp\left[N \frac{\Delta_{n+1} - D_n}{2D_n}\right] \qquad (2.2.5.29)$$

and

$$P(E_1, E_2, \dots, E_n)$$
$$= (2\pi)^{-n} D_n^{-1/2} \exp\left[N \frac{\Delta_{n+1} - D_n}{D_n}\right], \qquad (2.2.5.30)$$

respectively. Because D_n is a constant, the maximum values of the conditional joint probabilities (2.2.5.29) and (2.2.5.30) are obtained when Δ_{n+1} is a maximum. Thus the *maximum determinant rule* may be stated (Tsoucaris, 1970; Lajzérowicz & Lajzérowicz, 1966): among all sets of phases which are compatible with the inequality

$$\Delta_{n+1}(E_1, E_2, \dots, E_n) \geq 0$$

the most probable one is that which leads to a maximum value of Δ_{n+1}.

If only one phase, *i.e.* φ_q, is unknown whereas all other phases and moduli are known then (de Rango *et al.*, 1974; Podjarny *et al.*, 1976) for cs. crystals

$$P^\pm(E_q) \simeq 0.5 + 0.5 \, \tanh \left\{ \pm|E_q| \sum_{\substack{p=1 \\ p \neq q}}^{n} \Lambda_{pq} E_p \right\}, \qquad (2.2.5.31)$$

and for ncs. crystals

$$P(\varphi_q) = [2\pi I_0(G_q)]^{-1} \exp\{G_q \cos(\varphi_q - \theta_q)\}, \qquad (2.2.5.32)$$

where

$$G_q \exp(i\theta_q) = 2|E_q| \sum_{p \neq q=1}^{n} \Lambda_{pq} E_p.$$

Equations (2.2.5.31) and (2.2.5.32) generalize (2.2.5.11) and (2.2.5.7), respectively, and reduce to them for $n = 3$. Fourth-order determinantal formulae estimating triplet invariants in cs. and ncs. crystals, and making use of the entire data set, have recently been secured (Karle, 1979, 1980a).

Advantages, limitations and applications of determinantal formulae can be found in the literature (Heinermann *et al.*, 1979; de Rango *et al.*, 1975, 1985). Taylor *et al.* (1978) combined K–H

determinants with a magic-integer approach. The computing time, however, was larger than that required by standard computing techniques. The use of K–H matrices has been made faster and more effective by de Gelder *et al.* (1990) (see also de Gelder, 1992). They developed a phasing procedure (*CRUNCH*) which uses random phases as starting points for the maximization of the K–H determinants.

2.2.5.8. Algebraic relationships for structure seminvariants

According to the representations method (Giacovazzo, 1977a, 1980a,b):

(i) any s.s. Φ may be estimated *via* one or more s.i.'s $\{\psi\}$, whose values differ from Φ by a constant arising because of symmetry;

(ii) two types of s.s.'s exist, first-rank and second-rank s.s.'s, with different algebraic properties:

(iii) conditions characterizing s.s.'s of first rank for any space group may be expressed in terms of seminvariant moduli and seminvariantly associated vectors. For example, for all the space groups with point group 422 [Hauptman–Karle group $(h + k, l)$ $P(2, 2)$] the one-phase s.s.'s of first rank are characterized by

$$(h, k, l) \equiv 0 \bmod (2, 2, 0) \text{ or } (2, 0, 2) \text{ or } (0, 2, 2)$$
$$(h \pm k, l) \equiv 0 \bmod (0, 2) \text{ or } (2, 0).$$

The more general expressions for the s.s.'s of first rank are

(*a*) $\Phi = \varphi_{\mathbf{u}} = \varphi_{\mathbf{h}(I-R_\alpha)}$ for one-phase s.s.'s;

(*b*) $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2} = \varphi_{\mathbf{h}_1-\mathbf{h}_2 R_\beta} + \varphi_{\mathbf{h}_2-\mathbf{h}_1 R_\alpha}$ for two-phase s.s.'s;

(*c*) $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2} + \varphi_{\mathbf{u}_3} = \varphi_{\mathbf{h}_1-\mathbf{h}_2 R_\beta} + \varphi_{\mathbf{h}_2-\mathbf{h}_3 R_\gamma} + \varphi_{\mathbf{h}_3-\mathbf{h}_1 R_\alpha}$ for three-phase s.s.'s;

(*d*) $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2} + \varphi_{\mathbf{u}_3} + \varphi_{\mathbf{u}_4}$
$$= \varphi_{\mathbf{h}_1-\mathbf{h}_2 R_\beta} + \varphi_{\mathbf{h}_2-\mathbf{h}_3 R_\gamma} + \varphi_{\mathbf{h}_3-\mathbf{h}_4 R_\delta} + \varphi_{\mathbf{h}_4-\mathbf{h}_1 R_\alpha}$$

for four-phase s.s.'s; *etc.*

In other words:

(*a*) $\varphi_{\mathbf{u}}$ is an s.s. of first rank if at least one $\mathbf{h}$ and at least one rotation matrix R_α exist such that $\mathbf{u} = \mathbf{h}(I - R_\alpha)$. $\varphi_{\mathbf{u}}$ may be estimated *via* the special triplet invariants

$$\{\psi\} = \varphi_{\mathbf{u}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}R_\alpha}. \qquad (2.2.5.33)$$

The set $\{\psi\}$ is called the *first representation* of $\varphi_{\mathbf{u}}$.

(*b*) $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2}$ is an s.s. of first rank if at least two vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ and two rotation matrices R_α and R_β exist such that

$$\begin{cases} \mathbf{u}_1 = \mathbf{h}_1 - \mathbf{h}_2 R_\beta \\ \mathbf{u}_2 = \mathbf{h}_2 - \mathbf{h}_1 R_\alpha. \end{cases} \qquad (2.2.5.34)$$

Φ may then be estimated *via* the special quartet invariants

$$\{\psi\} = \varphi_{\mathbf{u}_1 R_\alpha} + \varphi_{\mathbf{u}_2} - \varphi_{\mathbf{h}_2} + \varphi_{\mathbf{h}_2 R_\beta R_\alpha} \qquad (2.2.5.35a)$$

and

$$\{\psi\} = \{\varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2 R_\beta} - \varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_1 R_\alpha R_\beta}\}. \qquad (2.2.5.35b)$$

For example, $\Phi = \varphi_{123} + \varphi_{\bar{7}\bar{2}\bar{5}}$ in $P2_1$ may be estimated *via*

$$\{\psi\} = \varphi_{123} + \varphi_{\bar{7}\bar{2}\bar{5}} - \varphi_{3K\bar{1}} + \varphi_{3K1}$$

and

$$\{\psi\} = \varphi_{123} + \varphi_{\bar{7}\bar{2}\bar{5}} - \varphi_{4K4} + \varphi_{\bar{4}K\bar{4}},$$

where K is a free index.

The set of special quartets (2.2.5.35a) and (2.2.5.35b) constitutes the *first representations* of Φ.

Structure seminvariants of the second rank can be characterized as follows: suppose that, for a given seminvariant Φ, it is not possible to find a vectorial index $\mathbf{h}$ and a rotation matrix R_α such that $\Phi - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}R_\alpha}$ is a structure invariant. Then Φ is a structure

seminvariant of the second rank and a set of structure invariants ψ can certainly be formed, of type

$$\{\psi\} = \Phi + \varphi_{\mathbf{h}R_p} - \varphi_{\mathbf{h}R_q} + \varphi_{\mathbf{l}R_i} - \varphi_{\mathbf{l}R_j},$$

by means of suitable indices $\mathbf{h}$ and $\mathbf{l}$ and rotation matrices R_p, R_q, R_i and R_j. As an example, for symmetry class 222, φ_{240} or φ_{024} or φ_{204} are s.s.'s of the first rank while φ_{246} is an s.s. of the second rank.

The procedure may easily be generalized to s.s.'s of any order of the first and of the second rank. So far only the role of one-phase and two-phase s.s.'s of the first rank in direct procedures is well documented (see references quoted in Sections 2.2.5.9 and 2.2.5.10).

2.2.5.9. Formulae estimating one-phase structure seminvariants of the first rank

Let $E_{\mathbf{H}}$ be our one-phase s.s. of the first rank, where

$$\mathbf{H} = \mathbf{h}(I - R_n). \qquad (2.2.5.36)$$

In general, more than one rotation matrix R_n and more than one vector $\mathbf{h}$ are compatible with (2.2.5.36). The set of special triplets

$$\{\psi\} = \{\varphi_{\mathbf{H}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}R_n}\}$$

is the first representation of $E_{\mathbf{H}}$. In cs. space groups the probability that $E_{\mathbf{H}} > 0$, given $|E_{\mathbf{H}}|$ and the set $\{|E_{\mathbf{h}}|\}$, may be estimated (Hauptman & Karle, 1953; Naya $et\ al.$, 1964; Cochran & Woolfson, 1955) by

$$P^+(E_{\mathbf{H}}) \simeq 0.5 + 0.5 \tanh \sum_{\mathbf{h}, n} G_{\mathbf{h}, n}(-1)^{2\mathbf{h}\cdot\mathbf{T}_n}, \qquad (2.2.5.37)$$

where

$$G_{\mathbf{h}, n} = |E_{\mathbf{H}}|\varepsilon_{\mathbf{h}}/(2\sqrt{N}), \ \text{ and } \ \varepsilon = |E|^2 - 1.$$

In (2.2.5.37), the summation over n goes within the set of matrices R_n for which (2.2.5.35a,b) is compatible, and $\mathbf{h}$ varies within the set of vectors which satisfy (2.2.5.36) for each R_n. Equation (2.2.5.36) is actually a generalized way of writing the so-called $\sum_1$ relationships (Hauptman & Karle, 1953).

If $\varphi_{\mathbf{H}}$ is a phase restricted by symmetry to $\theta_{\mathbf{H}}$ and $\theta_{\mathbf{H}} + \pi$ in an ncs. space group then (Giacovazzo, 1978)

$$P(\varphi_{\mathbf{H}} = \theta_{\mathbf{H}}) \simeq 0.5 + 0.5 \tanh\left\{\sum_{\mathbf{h}, n} G_{\mathbf{h}, n} \cos(\theta_{\mathbf{H}} - 2\pi\mathbf{h}\cdot\mathbf{T}_n)\right\}. \qquad (2.2.5.38)$$

If $\varphi_{\mathbf{H}}$ is a general phase then $\varphi_{\mathbf{H}}$ is distributed according to

$$P(\varphi_{\mathbf{H}}) \simeq \frac{1}{L}\exp\{\alpha\cos(\varphi_{\mathbf{H}} - \theta_{\mathbf{H}})\},$$

where

$$\tan\theta_{\mathbf{H}} = \frac{\left(\sum\limits_{\mathbf{h}, n} G_{\mathbf{h}, n}\sin 2\pi\mathbf{h}\cdot\mathbf{T}_n\right)}{\left(\sum\limits_{\mathbf{h}, n} G_{\mathbf{h}, n}\cos 2\pi\mathbf{h}\cdot\mathbf{T}_n\right)} \qquad (2.2.5.39)$$

with a reliability measured by

$$\alpha = \left\{\left(\sum_{\mathbf{h}, n} G_{\mathbf{h}, n}\sin 2\pi\mathbf{h}\cdot\mathbf{T}_n\right)^2 + \left(\sum_{\mathbf{h}, n} G_{\mathbf{h}, n}\cos 2\pi\mathbf{h}\cdot\mathbf{T}_n\right)^2\right\}^{1/2}.$$

The second representation of $\varphi_{\mathbf{H}}$ is the set of special quintets

$$\{\psi\} = \{\varphi_{\mathbf{H}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}R_n} + \varphi_{\mathbf{k}R_j} - \varphi_{\mathbf{k}R_j}\} \qquad (2.2.5.40)$$

provided that $\mathbf{h}$ and R_n vary over the vectors and matrices for which (2.2.5.36) is compatible, $\mathbf{k}$ over the asymmetric region of the reciprocal space, and R_j over the rotation matrices in the space group. Formulae estimating $\varphi_{\mathbf{H}}$ via the second representation in all the space groups [all the base and cross magnitudes of the quintets (2.2.5.40) now constitute the $a\ priori$ information] have been secured (Giacovazzo, 1978; Cascarano & Giacovazzo, 1983; Cascarano, Giacovazzo, Calabrese $et\ al.$, 1984). Such formulae contain, besides the contribution of order $N^{-1/2}$ provided by the first representation, a supplementary (not negligible) contribution of order $N^{-3/2}$ arising from quintets.

Denoting

$$E_1 = E_{\mathbf{H}}, \ E_2 = E_{\mathbf{h}}, \ E_3 = E_{\mathbf{k}},$$
$$E_{4, j} = E_{\mathbf{h}+\mathbf{k}R_j}, \ E_{5, j} = E_{\mathbf{H}+\mathbf{k}R_j},$$

formulae (2.2.5.37), (2.2.5.38), (2.2.5.39) still hold provided that $\sum_{\mathbf{h}, n} G_{\mathbf{h}, n}$ is replaced by

$$\sum_{\mathbf{h}, n} G_{\mathbf{h}, n} + \sum_{\mathbf{h}, \mathbf{k}, n}' \frac{|E_{\mathbf{H}}|}{2N^{3/2}} \frac{A_{\mathbf{h}, \mathbf{k}, n}}{1 + B_{\mathbf{h}, \mathbf{k}, n}},$$

where

$$A_{\mathbf{h}, \mathbf{k}, n} = \left[(2|E_2|^2 - 1)\varepsilon_3 \left(\sum_{\substack{R_j = R_i \\ R_j + R_iR_n = 0}} \varepsilon_{4, i}\varepsilon_{5, j} + \sum_{\substack{R_j = R_iR_n \\ R_i = R_jR_n}} \varepsilon_{4, i}\varepsilon_{4, j}\right) \right.$$
$$\left. - \frac{\varepsilon_3}{2}\sum_{j=1}^m \varepsilon_{4, j} - \frac{1}{2}\sum_{\substack{R_j = R_i \\ R_j + R_iR_n = 0}} \varepsilon_{4, i}\varepsilon_{5, j}\right]\bigg/N,$$

$$B_{\mathbf{h}, \mathbf{k}, n} = \left[\varepsilon_1\varepsilon_3\sum_{j=1}^m \varepsilon_{5, j} + \varepsilon_1\sum_{\substack{R_j = R_iR_n \\ R_i = R_jR_n}} \varepsilon_{4, i}\varepsilon_{4, j} + \varepsilon_2\varepsilon_3\sum_{j=1}^m \varepsilon_{4, j}\right.$$
$$\left. + \varepsilon_2\sum_{\substack{R_j = R_i \\ R_j + R_iR_n = 0}} \varepsilon_{4, i}\varepsilon_{5, j} + \frac{1}{4}\varepsilon_1 H_4(E_2)\right]\bigg/(2N).$$

m is the number of symmetry operators and $H_4(E) = E^4 - 6E^2 + 3$ is the Hermite polynomial of order four.

$B_{\mathbf{h}, \mathbf{k}, n}$ is assumed to be zero if it is computed negative. The prime to the summation warns the reader that precautions have to be taken in order to avoid duplication in the contributions.

2.2.5.10. Formulae estimating two-phase structure seminvariants of the first rank

Two-phase s.s.'s of the first rank were first evaluated in some cs. space groups by the method of coincidence by Grant $et\ al.$ (1957); the idea was extended to ncs. space groups by Debaerdemaeker & Woolfson (1972), and in a more general way by Giacovazzo (1977e,f).

The technique was based on the combination of the two triplets

$$\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} \simeq \varphi_{\mathbf{h}_1+\mathbf{h}_2}$$
$$\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2R} \simeq \varphi_{\mathbf{h}_1+\mathbf{h}_2R},$$

which, subtracted from one another, give

$$\varphi_{\mathbf{h}_1+\mathbf{h}_2R} - \varphi_{\mathbf{h}_1+\mathbf{h}_2} \simeq \varphi_{\mathbf{h}_2R} - \varphi_{\mathbf{h}_2} \simeq -2\pi\mathbf{h}\cdot\mathbf{T}.$$

If all four $|E|$'s are sufficiently large, an estimate of the two-phase seminvariant $\varphi_{\mathbf{h}_1+\mathbf{h}_2 R} - \varphi_{\mathbf{h}_1+\mathbf{h}_2}$ is available.

Probability distributions valid in $P2_1$ according to the neighbourhood principle have been given by Hauptman & Green (1978). Finally, the theory of representations was combined by Giacovazzo (1979a) with the joint probability distribution method in order to estimate two-phase s.s.'s in all the space groups.

According to representation theory, the problem is that of evaluating $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2}$ via the special quartets (2.2.5.35a) and (2.2.5.35b). Thus, contributions of order N^{-1} will appear in the probabilistic formulae, which will be functions of the basis and of the cross magnitudes of the quartets (2.2.5.35) . Since more pairs of matrices $\mathbf{R}_\alpha$ and $\mathbf{R}_\beta$ can be compatible with (2.2.5.34), and for each pair $(\mathbf{R}_\alpha, \mathbf{R}_\beta)$ more pairs of vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ may satisfy (2.2.5.34), several quartets can in general be exploited for estimating Φ. The simplest case occurs in $P\bar{1}$ where the two quartets (2.2.5.35) suggest the calculation of the six-variate distribution function $(\mathbf{u}_1 = \mathbf{h}_1 + \mathbf{h}_2, \mathbf{u}_2 = \mathbf{h}_1 - \mathbf{h}_2)$

$$P(E_{\mathbf{h}_1}, E_{\mathbf{h}_2}, E_{\mathbf{h}_1+\mathbf{h}_2}, E_{\mathbf{h}_1-\mathbf{h}_2}, E_{2\mathbf{h}_1}, E_{2\mathbf{h}_2})$$

which leads to the probability formula

$$P^+ \simeq 0.5 + 0.5 \tanh\left(\frac{|E_{\mathbf{h}_1+\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2}|}{2N} \cdot \frac{A}{1+B}\right),$$

where P^+ is the probability that the product $E_{\mathbf{h}_1+\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2}$ is positive, and

$$A = \varepsilon_{\mathbf{h}_1} + \varepsilon_{\mathbf{h}_2} + 2\varepsilon_{\mathbf{h}_1}\varepsilon_{\mathbf{h}_2} + \varepsilon_{\mathbf{h}_1}\varepsilon_{2\mathbf{h}_1} + \varepsilon_{\mathbf{h}_2}\varepsilon_{2\mathbf{h}_2}$$
$$B = (\varepsilon_{\mathbf{h}_1}\varepsilon_{\mathbf{h}_2}\varepsilon_{\mathbf{u}_1} + \varepsilon_{\mathbf{h}_1}\varepsilon_{\mathbf{h}_2}\varepsilon_{\mathbf{u}_2}$$
$$+ \varepsilon_{\mathbf{u}_1}\varepsilon_{\mathbf{u}_2}\varepsilon_{2\mathbf{h}_1} + \varepsilon_{\mathbf{u}_1}\varepsilon_{\mathbf{u}_2}\varepsilon_{2\mathbf{h}_2})/(2N).$$

It may be seen that in favourable cases $P^+ < 0.5$.

For the sake of brevity, the probabilistic formulae for the general case are not given and the reader is referred to the original papers.

2.2.6. Direct methods in real and reciprocal space: Sayre's equation

The statistical treatment suggested by Wilson for scaling observed intensities corresponds, in direct space, to the origin peak of the Patterson function, so it is not surprising that a general correspondence exists between probabilistic formulation in reciprocal space and algebraic properties in direct space.

For a structure containing atoms which are fully resolved from one another, the operation of raising $\rho(\mathbf{r})$ to the nth power retains the condition of resolved atoms but changes the shape of each atom. Let

$$\rho(\mathbf{r}) = \sum_{j=1}^{N} \rho_j(\mathbf{r} - \mathbf{r}_j),$$

where $\rho_j(\mathbf{r})$ is an atomic function and $\mathbf{r}_j$ is the coordinate of the 'centre' of the atom. Then the Fourier transform of the electron density can be written as

$$F_{\mathbf{h}} = \sum_{j=1}^{N} \int_V \rho_j(\mathbf{r} - \mathbf{r}_j) \exp(2\pi i \mathbf{h} \cdot \mathbf{r}) \, dV$$
$$= \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j). \qquad (2.2.6.1)$$

If the atoms do not overlap

$$\rho^n(\mathbf{r}) = \left[\sum_{j=1}^{N} \rho_j(\mathbf{r} - \mathbf{r}_j)\right]^n \simeq \sum_{j=1}^{N} \rho_j^n(\mathbf{r} - \mathbf{r}_j)$$

and its Fourier transform gives

$$_nF_{\mathbf{h}} = \int_V \rho^n(\mathbf{r}) \exp(2\pi i \mathbf{h} \cdot \mathbf{r}) \, dV$$
$$= \sum_{j=1}^{N} {}_nf_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j). \qquad (2.2.6.2)$$

$_nf_j$ is the scattering factor for the jth peak of $\rho^n(\mathbf{r})$:

$$_nf_j(\mathbf{h}) = \int_V \rho_j^n(\mathbf{r}) \exp(2\pi i \mathbf{h} \cdot \mathbf{r}) \, d\mathbf{r}.$$

We now introduce the condition that all atoms are equal, so that $f_j \equiv f$ and $_nf_j \equiv {}_nf$ for any j. From (2.2.6.1) and (2.2.6.2) we may write

$$F_{\mathbf{h}} = \frac{f}{_nf} {}_nF_{\mathbf{h}} = \theta_n \, {}_nF_{\mathbf{h}}, \qquad (2.2.6.3)$$

where θ_n is a function which corrects for the difference of shape of the atoms with electron distributions $\rho(\mathbf{r})$ and $\rho^n(\mathbf{r})$. Since

$$\rho^n(\mathbf{r}) = \rho(\mathbf{r}) \ldots \rho(\mathbf{r})$$
$$= \frac{1}{V^n} \sum_{\substack{\mathbf{h}_1, \ldots, \mathbf{h}_n \\ -\infty}}^{+\infty} F_{\mathbf{h}_1} \ldots F_{\mathbf{h}_n} \exp[-2\pi i(\mathbf{h}_1 + \ldots + \mathbf{h}_n) \cdot \mathbf{r}],$$

the Fourier transform of both sides gives

$$_nF_{\mathbf{h}} = \frac{1}{V^n} \sum_{\substack{\mathbf{h}_1, \ldots, \mathbf{h}_n \\ -\infty}}^{+\infty} F_{\mathbf{h}_1} \ldots F_{\mathbf{h}_n} \int_V \exp[2\pi i(\mathbf{h} - \mathbf{h}_1 - \ldots - \mathbf{h}_n) \cdot \mathbf{r}] \, dV$$
$$= \frac{1}{V^{n-1}} \sum_{\substack{\mathbf{h}_1, \ldots, \mathbf{h}_{n-1} \\ -\infty}}^{+\infty} F_{\mathbf{h}_1} F_{\mathbf{h}_2} \ldots F_{\mathbf{h}-\mathbf{h}_1-\mathbf{h}_2-\ldots-\mathbf{h}_{n-1}},$$

from which the following relation arises:

$$F_{\mathbf{h}} = \theta_n \frac{1}{V^{n-1}} \sum_{\substack{\mathbf{h}_1, \ldots, \mathbf{h}_{n-1} \\ -\infty}}^{+\infty} F_{\mathbf{h}_1} F_{\mathbf{h}_2} \ldots F_{\mathbf{h}-\mathbf{h}_1-\mathbf{h}_2-\ldots-\mathbf{h}_{n-1}}. \qquad (2.2.6.4)$$

For $n = 2$, equation (2.2.6.4) reduces to Sayre's (1952) equation [but see also Hughes (1953)]

$$F_{\mathbf{h}} = \theta_2 \frac{1}{V} \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}}. \qquad (2.2.6.5)$$

If the structure contains resolved isotropic atoms of two types, P and Q, it is impossible to find a factor θ_2 such that the relation $F_{\mathbf{h}} = \theta_2 \, {}_2F_{\mathbf{h}}$ holds, since this would imply values of θ_2 such that $(_2f)_P = \theta_2(f)_P$ and $(_2f)_Q = \theta_2(f)_Q$ simultaneously. However, the following relationship can be stated (Woolfson, 1958):

$$F_{\mathbf{h}} = \frac{A_s}{V} \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}} + \frac{B_s}{V^2} \sum_{\mathbf{k},\mathbf{l}} F_{\mathbf{k}} F_{\mathbf{l}} F_{\mathbf{h}-\mathbf{k}-\mathbf{l}}, \qquad (2.2.6.6)$$

where A_s and B_s are adjustable parameters of $(\sin\theta)/\lambda$. Equation (2.2.6.6) can easily be generalized to the case of structures containing resolved atoms of more than two types (von Eller, 1973).

Besides the algebraic properties of the electron density, Patterson methods also can be developed so that they provide phase indications. For example, it is possible to find the reciprocal counterpart of the function

$$P_n(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \int_V \rho(\mathbf{r})\rho(\mathbf{r} + \mathbf{u}_1) \ldots \rho(\mathbf{r} + \mathbf{u}_n) \, dV. \qquad (2.2.6.7)$$

For $n = 1$ the function (2.2.6.7) coincides with the usual Patterson function $P(\mathbf{u})$; for $n = 2$, (2.2.6.7) reduces to the double Patterson function $P_2(\mathbf{u}_1, \mathbf{u}_2)$ introduced by Sayre (1953). Expansion of $P_2(\mathbf{u}_1, \mathbf{u}_2)$ as a Fourier series yields

$$P_2(\mathbf{u}_1, \mathbf{u}_2) = \frac{1}{V^2} \sum_{\mathbf{h}_1, \mathbf{h}_2} E_{\mathbf{h}_1} E_{\mathbf{h}_2} E_{\mathbf{h}_3} \exp[-2\pi i(\mathbf{h}_1 \cdot \mathbf{u}_1 + \mathbf{h}_2 \cdot \mathbf{u}_2)].$$

$$(2.2.6.8)$$

Vice versa, the value of a triplet invariant may be considered as the Fourier transform of the double Patterson.

Among the main results relating direct- and reciprocal-space properties it may be remembered:

(*a*) from the properties of $P_2(\mathbf{u}_1, \mathbf{u}_2)$ the following relationship may be obtained (Vaughan, 1958)

$$E_{\mathbf{h}_1} E_{\mathbf{h}_2} E_{\mathbf{h}_1+\mathbf{h}_2} - N^{-3/2}$$
$$\simeq A_1 \langle (|E_{\mathbf{k}}|^2 - 1)(|E_{\mathbf{h}_1+\mathbf{k}}|^2 - 1)(|E_{-\mathbf{h}_2+\mathbf{k}}|^2 - 1) \rangle_{\mathbf{k}} - B_1,$$

which is clearly related to (2.2.5.12);

(*b*) the zero points in the Patterson function provide information about the value of a triplet invariant (Anzenhofer & Hoppe, 1962; Allegra, 1979);

(*c*) the Hoppe sections (Hoppe, 1963) of the double Patterson provide useful information for determining the triplet signs (Krabbendam & Kroon, 1971; Simonov & Weissberg, 1970);

(*d*) one phase s.s.'s of the first rank can be estimated *via* the Fourier transform of single Harker sections of the Patterson (Ardito *et al.*, 1985), *i.e.*

$$F_{\mathbf{H}} \sim \frac{1}{L} \exp(2\pi i \mathbf{h} \cdot \mathbf{T}_n) \int\limits_{HS(\mathbf{I}, \mathbf{C}_n)} P(\mathbf{u}) \exp(2\pi i \mathbf{h} \cdot \mathbf{u}) \, d\mathbf{u}, \quad (2.2.6.9)$$

where (see Section 2.2.5.9) $\mathbf{H} = \mathbf{h}(\mathbf{I} - \mathbf{R}_n)$ is the s.s., $\mathbf{u}$ varies over the complete Harker section corresponding to the operator $\mathbf{C}_n$ [in symbols $HS(\mathbf{I}, \mathbf{C}_n)$] and L is a constant which takes into account the dimensionality of the Harker section.

If no spurious peak is on the Harker section, then (2.2.6.9) is an exact relationship. Owing to the finiteness of experimental data and to the presence of spurious peaks, (2.2.6.9) cannot be considered in practice an exact relation: it works better when heavy atoms are in the chemical formula.

More recently (Cascarano, Giacovazzo, Luić *et al.*, 1987), a special least-squares procedure has been proposed for discriminating spurious peaks among those lying on Harker sections and for improving positional and thermal parameters of heavy atoms.

(*e*) translation and rotation functions (see Chapter 2.3), when defined in direct space, always have their counterpart in reciprocal space.

2.2.7. Scheme of procedure for phase determination: the small-molecule case

A traditional procedure for phase assignment may be schematically presented as follows:

Stage 1: *Normalization of s.f.'s.* See Section 2.2.4.

Stage 2: *(Possible) estimation of one-phase s.s.'s.* The computing program recognizes the one-phase s.s.'s and applies the proper formulae (see Section 2.2.5.9).

Each phase is associated with a reliability value, to allow the user to regard as known only those phases with reliability higher than a given threshold.

Stage 3: *Search of the triplets.* The reflections are listed for decreasing $|E|$ values and, related to each $|E|$ value, all possible triplets are reported (this is the so-called $\sum_2$ list). The value $G = 2|E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}|/\sqrt{N}$ is associated with every triplet for an evaluation of its efficiency. Usually reflections with $|E| < E_s$ (E_s may range from 1.2 to 1.6) are omitted from this stage onward.

Stage 4: *Definition of the origin and enantiomorph.* This stage is carried out according to the theory developed in Section 2.2.3. Phases chosen for defining the origin and enantiomorph, one-phase seminvariants estimated at stage 2, and symbolic phases described at stage 5 are the only phases known at the beginning of the phasing procedure. This set of phases is conventionally referred to as the *starting set*, from which iterative application of the tangent formula will derive new phase estimates.

Stage 5: *Assignment of one or more (symbolic or numerical) phases.* In complex structures the number of phases assigned for fixing the origin and the enantiomorph may be inadequate as a basis for further phase determination. Furthermore, only a few one-phase s.s.'s can be determined with sufficient reliability to make them qualify as members of the starting set. Symbolic phases may then be associated with some (generally from 1 to 6) high-modulus reflections (symbolic addition procedures). Iterative application of triplet relations leads to the determination of other phases which, in part, will remain expressed by symbols (Karle & Karle, 1966).

In other procedures (multisolution procedures) each symbol is assigned four phase values in turn: $\pi/4, 3\pi/4, 5\pi/4, 7\pi/4$. If p symbols are used, in at least one of the possible 4^p solutions each symbolic phase has unit probability of being within $45°$ of its true value, with a mean error of $22.5°$.

To find a good starting set a *convergence method* (Germain *et al.*, 1970) is used according to which: (*a*)

$$\langle \alpha_{\mathbf{h}} \rangle = \sum_j G_j I_1(G_j)/I_0(G_j)$$

is calculated for all reflections (j runs over the set of triplets containing $\mathbf{h}$); (*b*) the reflection is found with smallest $\langle \alpha \rangle$ not already in the starting set; it is retained to define the origin if the origin cannot be defined without it; (*c*) the reflection is eliminated if it is not used for origin definition. Its $\langle \alpha \rangle$ is recorded and $\langle \alpha \rangle$ values for other reflections are updated; (*d*) the cycle is repeated from (*b*) until all reflections are eliminated; (*e*) the reflections with the smallest $\langle \alpha \rangle$ at the time of elimination go into the starting set; (*f*) the cycle from (*a*) is repeated until all reflections have been chosen.

Stage 6: *Application of tangent formula.* Phases are determined in reverse order of elimination in the convergence procedure. In order to ensure that poorly determined phases $\varphi_{\mathbf{k}_j}$ and $\varphi_{\mathbf{h}-\mathbf{k}_j}$ have little effect in the determination of other phases a weighted tangent formula is normally used (Germain *et al.*, 1971):

$$\tan \varphi_{\mathbf{h}} = \frac{\sum_j w_{\mathbf{k}_j} w_{\mathbf{h}-\mathbf{k}_j} |E_{\mathbf{k}_j} E_{\mathbf{h}-\mathbf{k}_j}| \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}{\sum_j w_{\mathbf{k}_j} w_{\mathbf{h}-\mathbf{k}_j} |E_{\mathbf{k}_j} E_{\mathbf{h}-\mathbf{k}_j}| \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}, \quad (2.2.7.1)$$

where

$$w_{\mathbf{h}} = \min (0.2\alpha, 1).$$

Once a large number of contributions are available in (2.2.7.1) for a given $\varphi_{\mathbf{h}}$, then the value of $\alpha_{\mathbf{h}}$ quickly becomes greater than 5, and so assigns an unrealistic unitary weight to $\varphi_{\mathbf{h}}$. In this respect a different weighting scheme may be proposed (Hull & Irwin, 1978) according to which

$$w = \psi \exp(-x^2) \int\limits_0^x \exp(t^2) \, dt, \quad (2.2.7.2)$$

where $x = \alpha/\langle \alpha \rangle$ and $\psi = 1.8585$ is a constant chosen so that $w = 1$ when $x = 1$. Except for ψ, the right-hand side of (2.2.7.2) is the Dawson integral which assumes its maximum value at $x = 1$ (see Fig. 2.2.7.1): when $\alpha > \langle \alpha \rangle$ or $\alpha < \langle \alpha \rangle$ then $w < 1$ and so the agreement between α and $\langle \alpha \rangle$ is promoted.

Alternative weighting schemes for the tangent formula are frequently used [for example, see Debaerdemaeker *et al.* (1985)]. In one (Giacovazzo, 1979*b*), the values $\alpha_{\mathbf{k}_j}$ and $\alpha_{\mathbf{h}-\mathbf{k}_j}$ (which are
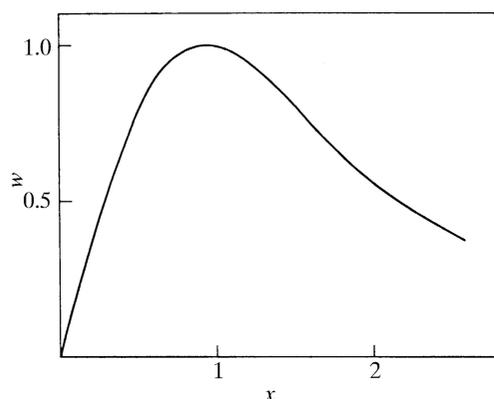
Fig. 2.2.7.1. The form of w as given by (2.2.7.2).

usually available in direct procedures) are considered as additional *a priori* information so that (2.2.7.1) may be replaced by

$$\tan \varphi_{\mathbf{h}} \simeq \frac{\sum_j \beta_j \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}{\sum_j \beta_j \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}, \qquad (2.2.7.3)$$

where β_j is the solution of the equation

$$D_1(\beta_j) = D_1(G_j)D_1(\alpha_{\mathbf{k}_j})D_1(\alpha_{\mathbf{h}-\mathbf{k}_j}). \qquad (2.2.7.4)$$

In (2.2.7.4),

$$G_j = 2|E_{\mathbf{h}}E_{\mathbf{k}_j}E_{\mathbf{h}-\mathbf{k}_j}|/\sqrt{N}$$

or the corresponding second representation parameter, and $D_1(x) = I_1(x)/I_0(x)$ is the ratio of two modified Bessel functions.

In order to promote (in accordance with the aims of Hull and Irwin) the agreement between α and $\langle \alpha \rangle$, the distribution of α may be used (Cascarano, Giacovazzo, Burla *et al.*, 1984; Burla *et al.*, 1987); in particular, the first two moments of the distribution: accordingly,

$$w = \left\{ \exp\left[\frac{-(\alpha - \langle\alpha\rangle)^2}{2\sigma_\alpha^2} \right] \right\}^{1/3}$$

may be used, where σ_α^2 is the estimated variance of α.

Stage 7: *Figures of merit.* The correct solution is found among several by means of figures of merit (FOMs) which are expected to be extreme for the correct solution. Largely used are (Germain *et al.*, 1970)

$$(a) \qquad \text{ABSFOM} = \sum_{\mathbf{h}} \alpha_{\mathbf{h}} / \sum_{\mathbf{h}} \langle \alpha_{\mathbf{h}} \rangle,$$

which is expected to be unity for the correct solution.

$$(b) \qquad \text{PSI0} = \frac{\sum_{\mathbf{h}} \left| \sum_{\mathbf{k}} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}} \right|}{\sum_{\mathbf{h}} \left(\sum_{\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}|^2 \right)^{1/2}}.$$

The summation over $\mathbf{k}$ includes (Cochran & Douglas, 1957) the strong $|E|$'s for which phases have been determined, and indices $\mathbf{h}$ correspond to very small $|E_{\mathbf{h}}|$. Minimal values of PSI0 (≤ 1.20) are expected to be associated with the correct solution.

$$(c) \qquad R_\alpha = \frac{\sum_{\mathbf{h}} |\alpha_{\mathbf{h}} - \langle \alpha_{\mathbf{h}} \rangle|}{\sum_{\mathbf{h}} \langle \alpha_{\mathbf{h}} \rangle}.$$

That is, the Karle & Karle (1966) residual between the actual and the estimated α's. After scaling of $\alpha_{\mathbf{h}}$ on $\langle \alpha_{\mathbf{h}} \rangle$ the correct solution should be characterized by the smallest R_α values.

$$(d) \qquad \text{NQEST} = \sum_j G_j \cos \Phi_j,$$

where G is defined by (2.2.5.21) and

$$\Phi = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{l}} - \varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}}$$

are quartet invariants characterized by large basis magnitudes and small cross magnitudes (De Titta *et al.*, 1975; Giacovazzo, 1976). Since G is expected to be negative as well as $\cos \Phi$, the value of NQEST is expected to be positive and a maximum for the correct solution.

Figures of merit are then combined as

$$\begin{aligned}
\text{CFOM} = {}& w_1 \frac{\text{ABSFOM} - \text{ABSFOM}_{\min}}{\text{ABSFOM}_{\max} - \text{ABSFOM}_{\min}} \\
& + w_2 \frac{\text{PSI0}_{\max} - \text{PSI0}}{\text{PSI0}_{\max} - \text{PSI0}_{\min}} \\
& + w_3 \frac{R_{\alpha_{\max}} - R_\alpha}{R_{\alpha_{\max}} - R_{\alpha_{\min}}} \\
& + w_4 \frac{\text{NQEST} - \text{NQEST}_{\min}}{\text{NQEST}_{\max} - \text{NQEST}_{\min}},
\end{aligned}$$

where w_i are empirical weights proportional to the confidence of the user in the various FOMs.

Different FOMs are often used by some authors in combination with those described above: for example, enantiomorph triplets and quartets are supplementary FOMs (Van der Putten & Schenk, 1977; Cascarano, Giacovazzo & Viterbo, 1987).

Different schemes of calculating and combining FOMs are also used: one scheme (Cascarano, Giacovazzo & Viterbo, 1987) uses

$$(a1) \qquad \text{CPHASE} = \frac{\sum w_j G_j \cos(\Phi_j - \theta_j) + w_j G_j \cos \Phi_j}{\sum_{\text{s.i.+s.s.}} w_j G_j D_1(G_j)},$$

where the first summation in the numerator extends over symmetry-restricted one-phase and two-phase s.s.'s (see Sections 2.2.5.9 and 2.2.5.10), and the second summation in the numerator extends over negative triplets estimated *via* the second representation formula [equation (2.2.5.13)] and over negative quartets. The value of CPHASE is expected to be close to unity for the correct solution.

$(a2)$ $\alpha_{\mathbf{h}}$ for strong triplets and $E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}$ contributions for PSI0 triplets may be considered random variables: the agreements between their actual and their expected distributions are considered as criteria for identifying the correct solution.

$(a3)$ correlation among some FOMs is taken into account.

According to this scheme, each FOM (as well as the CFOM) is expected to be unity for the correct solution. Thus one or more figures are available which constitute a sort of criterion (on an absolute scale) concerning the correctness of the various solutions: FOMs (and CFOM) $\simeq 1$ probably denote correct solutions, CFOMs $\ll 1$ should indicate incorrect solutions.

Stage 8: *Interpretation of E maps.* This is carried out in up to four stages (Koch, 1974; Main & Hull, 1978; Declercq *et al.*, 1973):

  (*a*) peak search;

  (*b*) separation of peaks into potentially bonded clusters;

  (*c*) application of stereochemical criteria to identify possible molecular fragments;

  (*d*) comparison of the fragments with the expected molecular structure.

2.2.8. Other multisolution methods applied to small molecules

In very complex structures a large initial set of known phases seems to be a basic requirement for a structure to be determined.

Table 2.2.8.1. *Magic-integer sequences for small numbers of phases (n) together with the number of sets produced and the root-mean-square error in the phases*

n	Sequence								No. of sets	R.m.s. error (°)
1	1								4	26
2	2	3							12	29
3	3	4	5						20	37
4	5	7	8	9					32	42
5	8	11	13	14	15				50	45
6	13	18	21	23	24	25			80	47
7	21	29	34	37	39	40	41		128	48
8	34	47	55	60	63	65	66	67	206	49

This aim can be achieved, for example, by introducing a large number of permutable phases into the initial set. However, the introduction of every new symbol implies a fourfold increase in computing time, which, even in fast computers, quickly leads to computing-time limitations. On the other hand, a relatively large starting set is not in itself enough to ensure a successful structure determination. This is the case, for example, when the triplet invariants used in the initial steps differ significantly from zero. New strategies have therefore been devised to solve more complex structures.

(1) *Magic-integer methods*

In the classical procedure described in Section 2.2.7, the unknown phases in the starting set are assigned all combinations of the values $\pm\pi/4$, $\pm3\pi/4$. For n unknown phases in the starting set, 4^n sets of phases arise by quadrant permutation; this is a number that increases very rapidly with n. According to White & Woolfson (1975), phases can be represented for a sequence of n integers by the equations

$$\varphi_i = m_i x \ (\text{mod } 2\pi), \quad i = 1, \ldots, n. \tag{2.2.8.1}$$

The set of equations can be regarded as the parametric equation of a straight line in n-dimensional phase space. The nature and size of errors connected with magic-integer representations have been investigated by Main (1977) who also gave a recipe for deriving magic-integer sequences which minimize the r.m.s. errors in the represented phases (see Table 2.2.8.1). To assign a phase value, the variable x in equation (2.2.8.1) is given a series of values at equal intervals in the range $0 < x < 2\pi$. The enantiomorph is defined by exploring only the appropriate half of the n-dimensional space.

A different way of using the magic-integer method (Declercq *et al.*, 1975) is the *primary–secondary P–S method* which may be described schematically in the following way:

(*a*) Origin- and enantiomorph-fixing phases are chosen and some one-phase s.s.'s are estimated.

(*b*) Nine phases [this is only an example: very long magic-integer sequences may be used to represent primary phases (Hull *et al.*, 1981; Debaerdemaeker & Woolfson, 1983)] are represented with the approximated relationships:

$$\begin{cases} \varphi_{i_1} = 3x \\ \varphi_{i_2} = 4x \\ \varphi_{i_3} = 5x \end{cases} \quad \begin{cases} \varphi_{j_1} = 3y \\ \varphi_{j_2} = 4y \\ \varphi_{j_3} = 5y \end{cases} \quad \begin{cases} \varphi_{p_1} = 3z \\ \varphi_{p_2} = 4z \\ \varphi_{p_3} = 5z. \end{cases}$$

Phases in (*a*) and (*b*) consititute the *primary set*.

(*c*) The phases in the *secondary set* are those defined through $\sum_2$ relationships involving pairs of phases from the primary set: they, too, can be expressed in magic-integer form.

(*d*) All the triplets that link together the phases in the combined primary and secondary set are now found, other than triplets used to obtain secondary reflections from the primary ones. The general algebraic form of these triplets will be

$$m_1 x + m_2 y + m_3 z + b \equiv 0 \ (\text{mod } 1),$$

where b is a phase constant which arises from symmetry translation. It may be expected that the 'best' value of the unknown x, y, z corresponds to a maximum of the function

$$\psi(x, y, z) = \sum |E_1 E_2 E_3| \cos 2\pi(m_1 x + m_2 y + m_3 z + b),$$

with $0 \le x, y, z < 1$. It should be noticed that ψ is a Fourier summation which can easily be evaluated. In fact, ψ is essentially a figure of merit for a large number of phases evaluated in terms of a small number of magic-integer variables and gives a measure of the internal consistency of $\sum_2$ relationships. The ψ map generally presents several peaks and therefore can provide several solutions for the variables.

(2) *The random-start method*

These are procedures which try to solve crystal structures by starting from random initial phases (Baggio *et al.*, 1978; Yao, 1981). They may be so described:

(*a*) A number of reflections (say NUM $\sim$ 100 or larger) at the bottom of the CONVERGE map are selected. These, and the relationships which link them, form the system for which trial phases will be found.

(*b*) A pseudo-random number generator is used to generate M sets of NUM random phases. Each of the M sets is refined and extended by the tangent formula or similar methods.

(3) *Accurate calculation of s.i.'s and s.s.'s with* 1, 2, 3, 4, ..., n *phases*

Having a large set of good phase relationships allows one to overcome difficulties in the early stages and in the refinement process of the phasing procedure. Accurate estimates of s.i.'s and s.s.'s may be achieved by the application of techniques such as the representation method or the neighbourhood principle (Hauptman, 1975; Giacovazzo, 1977*a*, 1980*b*). So far, second-representation formulae are available for triplets and one-phase seminvariants; in particular, reliably estimated negative triplets can be recognized, which is of great help in the phasing process (Cascarano, Giacovazzo, Camalli *et al.*, 1984). Estimation of higher-order s.s.'s with upper representations or upper neighbourhoods is rather difficult, both because the procedures are time consuming and because the efficiency of the present joint probability distribution techniques deteriorates with complexity. However, further progress can be expected in the field.

(4) *Modified tangent formulae and least-squares determination and refinement of phases*

The problem of deriving the individual phase angles from triplet relationships is greatly overdetermined: indeed the number of triplets, in fact, greatly exceeds the number of phases so that any $\varphi_\mathbf{h}$ may be determined by a least-squares approach (Hauptman *et al.*, 1969). The function to be minimized may be

$$M = \frac{\sum_\mathbf{k} w_\mathbf{k} [\cos(\varphi_\mathbf{h} - \varphi_\mathbf{k} - \varphi_{\mathbf{h}-\mathbf{k}}) - C_\mathbf{k}]^2}{\sum w_\mathbf{k}},$$

where $C_\mathbf{k}$ is the estimate of the cosine obtained by probabilistic or other methods.

Effective least-squares procedures based on linear equations (Debaerdemaeker & Woolfson, 1983; Woolfson, 1977) can also be used. A triplet relationship is usually represented by

$$(\varphi_p \pm \varphi_q \pm \varphi_r + b) \approx 0 \ (\text{mod } 2\pi), \tag{2.2.8.2}$$

where b is a factor arising from translational symmetry. If (2.2.8.2) is expressed in cycles and suitably weighted, then it may be written as

$$w(\varphi_p \pm \varphi_q \pm \varphi_r + b) = wn,$$

where n is some integer. If the integers were known then the equation would appear (in matrix notation) as

$$A\Phi = C, \qquad (2.2.8.3)$$

giving the least-squares solution

$$\Phi = (A^T A)^{-1} A^T C. \qquad (2.2.8.4)$$

When approximate phases are available, the nearest integers may be found and equations (2.2.8.3) and (2.2.8.4) constitute the basis for further refinement.

Modified tangent procedures are also used, such as (Sint & Schenk, 1975; Busetta, 1976)

$$\tan \varphi_{\mathbf{h}} \simeq \frac{\sum_j G_{\mathbf{h}, \mathbf{k}_j} \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j} - \Delta_j)}{\sum G_{\mathbf{h}, \mathbf{k}_j} \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j} - \Delta_j)},$$

where Δ_j is an estimate for the triplet phase sum $(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}_j} - \varphi_{\mathbf{h}-\mathbf{k}_j})$.

(5) *Techniques based on the positivity of Karle–Hauptman determinants*

(The main formulae have been briefly described in Section 2.2.5.7.) The maximum determinant rule has been applied to solve small structures (de Rango, 1969; Vermin & de Graaff, 1978) *via* determinants of small order. It has, however, been found that their use (Taylor *et al.*, 1978) is not of sufficient power to justify the larger amount of computing time required by the technique as compared to that required by the tangent formula.

(6) *Tangent techniques using simultaneously triplets, quartets, . . .*

The availability of a large number of phase relationships, in particular during the first stages of a direct procedure, makes the phasing process easier. However, quartets are sums of two triplets with a common reflection. If the phase of this reflection (and/or of the other cross terms) is known then the quartet probability formulae described in Section 2.2.5.5 cannot hold. Similar considerations may be made for quintet relationships. Thus triplet, quartet and quintet formulae described in the preceding paragraphs, if used without modifications, will certainly introduce systematic errors in the tangent refinement process.

A method which takes into account correlation between triplets and quartets has been described (Giacovazzo, 1980c) [see also Freer & Gilmore (1980) for a first application], according to which

$$\tan \varphi_{\mathbf{h}} \simeq \frac{\sum_{\mathbf{k}} G \sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) - \sum_{\mathbf{k},\mathbf{l}} G' \sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}})}{\sum_{\mathbf{k}} G \cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) - \sum_{\mathbf{k},\mathbf{l}} G' \cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}})},$$

where G' takes into account both the magnitudes of the cross terms of the quartet and the fact that their phases may be known.

(7) *Integration of Patterson techniques and direct methods (Egert & Sheldrick, 1985) [see also Egert (1983, and references therein)]*

A fragment of known geometry is oriented in the unit cell by real-space Patterson rotation search (see Chapter 2.3) and its position is found by application of a translation function (see Section 2.2.5.4 and Chapter 2.3) or by maximizing the weighted sum of the cosines of a small number of strong translation-sensitive triple phase invariants, starting from random positions. Suitable FOMs rank the most reliable solutions.

(8) *Maximum entropy methods*

A common starting point for all direct methods is a stochastic process according to which crystal structures are thought of as being generated by randomly placing atoms in the asymmetric unit of the unit cell according to some *a priori* distribution. A non-uniform prior distribution of atoms $p(\mathbf{r})$ gives rise to a source of random atomic positions with entropy (Jaynes, 1957)

$$H(p) = -\int_V p(\mathbf{r}) \log p(\mathbf{r}) \, d\mathbf{r}.$$

The maximum value $H_{\max} = \log V$ is reached for a uniform prior $p(\mathbf{r}) = 1/V$.

The strength of the restrictions introduced by $p(\mathbf{r})$ is not measured by $H(p)$ but by $H(p) - H_{\max}$, given by

$$H(p) - H_{\max} = -\int_V p(\mathbf{r}) \log[\,p(\mathbf{r})/m(\mathbf{r})] \, d\mathbf{r},$$

where $m(\mathbf{r}) = 1/V$. Accordingly, if a prior prejudice $m(\mathbf{r})$ exists, which maximizes H, the revised relative entropy is

$$S(p) = -\int_V p(\mathbf{r}) \log[\,p(\mathbf{r})/m(\mathbf{r})] \, d\mathbf{r}.$$

The maximization problem was solved by Jaynes (1957). If $G_j(p)$ are linear constraint functionals defined by given constraint functions $C_j(\mathbf{r})$ and constraint values c_j, *i.e.*

$$G_j(p) = \int_V p(\mathbf{r}) C_j(\mathbf{r}) \, d\mathbf{r} = c_j,$$

the most unbiased probability density $p(\mathbf{r})$ under prior prejudice $m(\mathbf{r})$ is obtained by maximizing the entropy of $p(\mathbf{r})$ relative to $m(\mathbf{r})$. A standard variational technique suggests that the constrained maximization is equivalent to the unconstrained maximization of the functional

$$S(p) + \sum_j \lambda_j G_j(p),$$

where the λ_j's are Lagrange multipliers whose values can be determined from the constraints.

Such a technique has been applied to the problem of finding good electron-density maps in different ways by various authors (Wilkins *et al.*, 1983; Bricogne, 1984; Navaza, 1985; Navaza *et al.*, 1983).

Maximum entropy methods are strictly connected with traditional direct methods: in particular it has been shown that:

(*a*) the maximum determinant rule (see Section 2.2.5.7) is strictly connected (Britten & Collins, 1982; Piro, 1983; Narayan & Nityananda, 1982; Bricogne, 1984);

(*b*) the construction of conditional probability distributions of structure factors amounts precisely to a reciprocal-space evaluation of the entropy functional $S(p)$ (Bricogne, 1984).

Maximum entropy methods are under strong development: important contributions can be expected in the near future even if a multipurpose robust program has not yet been written.

2.2.9. Some references to direct-methods packages: the small-molecule case

Some references for direct-methods packages are given below. Other useful packages using symbolic addition or multisolution procedures do exist but are not well documented.

CRUNCH: Gelder, R. de, de Graaff, R. A. G. & Schenk, H. (1993). *Automatic determination of crystal structures using Karle–Hauptman matrices. Acta Cryst.* A**49**, 287–293.

DIRDIF: Beurskens, P. T., Beurskens G., de Gelder, R., Garcia-Granda, S., Gould, R. O., Israel, R. & Smits, J. M. M. (1999). *The DIRDIF-99 program system.* Crystallography Laboratory, University of Nijmegen, The Netherlands.

MITHRIL: Gilmore, C. J. (1984). *MITHRIL. An integrated direct-methods computer program. J. Appl. Cryst.* **17**, 42–46.

MULTAN88: Main, P., Fiske, S. J., Germain, G., Hull, S. E., Declercq, J.-P., Lessinger, L. & Woolfson, M. M. (1999). *Crystallographic software: teXsan for Windows.* http://www.rigaku.com/downloads/journal/Vol15.1.1998/texsan.pdf.

PATSEE: Egert, E. & Sheldrick, G. M. (1985). *Search for a fragment of known geometry by integrated Patterson and direct methods. Acta Cryst.* A**41**, 262–268.

SAPI: Fan, H.-F. (1999). *Crystallographic software: teXsan for Windows.* http://www.rigaku.com/downloads/journal/Vol15.1.1998/texsan.pdf.

SnB: Weeks, C. M. & Miller, R. (1999). *The design and implementation of SnB version 2.0. J. Appl. Cryst.* **32**, 120–124.

SHELX97 and *SHELXS*: Sheldrick, G. M. (2000). *The SHELX home page.* http://shelx.uni-ac.gwdg.de/SHELX/.

SHELXD: Sheldrick, G. M. (1998). *SHELX: applications to macromolecules.* In *Direct methods for solving macromolecular structures*, edited by S. Fortier, pp. 401–411. Dordrecht: Kluwer Academic Publishers.

SIR97: Altomare, A., Burla, M. C., Camalli, M., Cascarano, G. L., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Polidori, G. & Spagna, R. (1999). *SIR97: a new tool for crystal structure determination and refinement. J. Appl. Cryst.* **32**, 115–119.

SIR2004: Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Polidori, G. & Spagna, R. (2005). *SIR2004: an improved tool for crystal structure determination and refinement. J. Appl. Cryst.* **38**, 381–388.

XTAL3.6.1: Hall, S. R., du Boulay, D. J. & Olthof-Hazekamp, R. (1999). *Xtal3.6 crystallographic software.* http://xtal.sourceforge.net/.

2.2.10. Direct methods in macromolecular crystallography

2.2.10.1. Introduction

The smallest protein molecules contain about 400 non-hydrogen atoms, so they cannot be solved *ab initio* by the algorithms specified in Sections 2.2.7 and 2.2.8. However, traditional direct methods are applied for:

(*a*) improvement of the accuracy of the available phases (refinement process);

(*b*) extension of phases from lower to higher resolution (phase-extension process).

The application of standard tangent techniques to (*a*) and (*b*) has not been found to be very satisfactory (Coulter & Dewar, 1971; Hendrickson *et al.*, 1973; Weinzierl *et al.*, 1969). Tangent methods, in fact, require atomicity and non-negativity of the electron density. Both these properties are not satisfied if data do not extend to atomic resolution ($d > 1.2$ Å). Because of series termination and other errors the electron-density map at $d > 1.2$ Å presents large negative regions which will appear as false peaks in the squared structure. However, tangent methods use only a part of the information given by the Sayre equation (2.2.6.5). In fact, (2.2.6.5) express two equations relating the radial and angular parts of the two sides, so obtaining a large degree of overdetermination of the phases. To achieve this Sayre (1972) [see also Sayre & Toupin (1975)] suggested minimizing (2.2.10.1) by least squares as a function of the phases:

$$\sum_{\mathbf{h}} \left| a_{\mathbf{h}} F_{\mathbf{h}} - \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}} \right|^2. \qquad (2.2.10.1)$$

Even if tests on rubredoxin (extensions of phases from 2.5 to 1.5 Å resolution) and insulin (Cutfield *et al.*, 1975) (from 1.9 to 1.5 Å resolution) were successful, the limitations of the method are its high cost and, especially, the higher efficiency of the least-squares method. Equivalent considerations hold for the application of determinantal methods to proteins [see Podjarny *et al.* (1981); de Rango *et al.* (1985) and literature cited therein].

A question now arises: why is the tangent formula unable to solve protein structures? Fan *et al.* (1991) considered the question from a first-principle approach and concluded that:

(1) the triplet phase probability distribution is very flat for proteins (N is very large) and close to the uniform distribution;

(2) low-resolution data create additional problems for direct methods since the number of available phase relationships per reflection is small.

Sheldrick (1990) suggested that direct methods are not expected to succeed if fewer than half of the reflections in the range 1.1–1.2 Å are observed with $|F| > 4\sigma(|F|)$ (a condition seldom satisfied by protein data).

The most complete analysis of the problem has been made by Giacovazzo, Guagliardi *et al.* (1994). They observed that the expected value of α (see Section 2.2.7) suggested by the tangent formula for proteins is comparable with the variance of the α parameter. In other words, for proteins the signal determining the phase is comparable with the noise, and therefore the phase indication is expected to be unreliable.

Quite relevant results have recently been obtained by integrating direct methods with some additional experimental information. In particular, we will describe the combination of direct methods with:

(*a*) direct-space techniques for the *ab initio* crystal structure solution of proteins;

(*b*) isomorphous-replacement (SIR–MIR) techniques;

(*c*) anomalous-dispersion (SAD–MAD) techniques;

(*d*) molecular replacement.

Point (*d*) will not be treated here, as it is described extensively in *IT* F, Part 13.

2.2.10.2. *Ab initio crystal structure solution of proteins*

Ab initio techniques do not require prior information of any atomic positions. The recent tremendous increase in computing speed led to direct methods evolving towards the rapid development of multisolution techniques. The new algorithms of the program *Shake-and-Bake* (Weeks *et al.*, 1994; Weeks & Miller, 1999; Hauptman *et al.*, 1999) allowed an impressive extension of the structural complexity amenable to direct phasing. In particular we mention: (*a*) the minimal principle (De Titta *et al.*, 1994), according to which the phase problem is considered as a constrained global optimization problem; (*b*) the refinement procedure, which alternately uses direct- and reciprocal-space techniques; and (*c*) the parameter-shift optimization technique (Bhuiya & Stanley, 1963), which aims at reducing the value of the minimal function (Hauptman, 1991; De Titta *et al.*, 1994). An effective variant of *Shake-and Bake* is *SHELXD* (Sheldrick, 1998) which cyclically alternates tangent refinement in reciprocal space with peak-list optimisation procedures in real space (Sheldrick & Gould, 1995). Detailed information on these programs is available in *IT* F (2001), Part 16.

A different approach is used by *ACORN* (Foadi *et al.*, 2000), which first locates a small fragment of the molecule (eventually by molecular-replacement techniques) to obtain a useful nonrandom starting set of phases, and then refines them by means of solvent-flattening techniques.

The program *SIR2004* (Burla *et al.*, 2005) uses the tangent formula as well as automatic Patterson techniques to obtain a first imperfect structural model; then direct-space techniques are used to refine the model. The Patterson approach is based on the use of the superposition minimum function (Buerger, 1959; Richardson & Jacobson, 1987; Sheldrick, 1992; Pavelčík, 1988; Pavelčík *et al.*, 1992; Burla *et al.*, 2004). It may be worth noting that even this approach is of multisolution type: up to 20 trial solutions are provided by using as pivots the highest maxima in the superposition minimum function.

It is today possible to solve structures up to 2500 non-hydrogen atoms in the asymmetric unit provided data at atomic (about 1 Å) resolution are available. Proteins with data at quasi-atomic resolution (say up to 1.5–1.6 Å) can also be solved, but with greater difficulties (Burla *et al.*, 2005). A simple evaluation of the potential of the *ab initio* techniques suggests that the structural complexity range and the resolution limits amenable to the *ab*

initio approach could be larger in the near future. The approach will profit by general technical advances like the increasing speed of computers and by the greater efficiency of informatic tools (*e.g.* faster Fourier-transform techniques). It could also profit from new specific crystallographic algorithms (for example, Oszlányi & Süto, 2004). It is of particular interest that extrapolating moduli and phases of nonmeasured reflections beyond the experimental resolution limit makes the *ab initio* phasing process more efficient, and leads to crystal structure solution even in cases in which the standard programs do not succeed (Caliandro *et al.*, 2005a,b). Moreover, the use of the extrapolated values improves the quality of the final electron-density maps and makes it easier to recognize the correct one among several trial structures.

2.2.10.3. Integration of direct methods with isomorphous replacement techniques

SIR–MIR cases are characterized by a situation in which there is one native protein and one or more heavy-atom substructures. In this situation the phasing procedure may be a two-step process: in the first stage the heavy-atom positions are identified by Patterson techniques (Rossmann, 1961; Okaya *et al.*, 1955) or by direct methods (Mukherjee *et al.*, 1989). In the second step the protein phases are estimated by exploiting the substructure information. Direct methods are able to contribute to both steps (see Sections 2.2.10.5 and 2.2.10.6). In Section 2.2.10.4 we show that direct methods are also able to suggest alternative one-step procedures by estimating structure invariants from isomorphous data.

2.2.10.4. SIR–MIR case: one-step procedures

The theoretical basis was established by Hauptman (1982a): his primary interest was to establish the two-phase and three-phase structure invariants by exploiting the experimental information provided by isomorphous data. The protein phases could be directly assigned *via* a tangent procedure.

Let us denote the modulus of the isomorphous difference as

$$\Delta F = |F_{\mathrm{d}}| - |F_{\mathrm{p}}|$$

where the subscripts d and p denote the derivative and the protein, respectively.

Denote also by f_j and g_j atomic scattering factors for the atom labelled j in a pair of isomorphous structures, and let $E_{\mathbf{h}}$ and $G_{\mathbf{h}}$ denote corresponding normalized structure factors. Then

$$E_{\mathbf{h}} = |E_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}) = \alpha_{20}^{-1/2} \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j),$$

$$G_{\mathbf{h}} = |G_{\mathbf{h}}| \exp(i\psi_{\mathbf{h}}) = \alpha_{02}^{-1/2} \sum_{j=1}^{N} g_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j),$$

where

$$\alpha_{mn} = \sum_{j=1}^{N} f_j^m g_j^n.$$

The conditional probability of the two-phase structure invariant $\Phi = \varphi_{\mathbf{h}} - \psi_{\mathbf{h}}$ given $|E_{\mathbf{h}}|$ and $|G_{\mathbf{h}}|$ is (Hauptman, 1982a)

$$P(\Phi \big| |E|, |G|) \simeq [2\pi I_0(Q)]^{-1} \exp(Q \cos \Phi),$$

where

$$Q = |EG|[2\alpha/(1 - \alpha^2)],$$
$$\alpha = \alpha_{11}/(\alpha_{20}^{1/2} \alpha_{02}^{1/2}).$$

Three-phase structure invariants were evaluated by considering that eight invariants exist for a given triple of indices **h**, **k**, **l** ($\mathbf{h} + \mathbf{k} + \mathbf{l} = 0$):

$$\Phi_1 = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} \qquad \Phi_2 = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \psi_{\mathbf{l}}$$
$$\Phi_3 = \varphi_{\mathbf{h}} + \psi_{\mathbf{k}} + \varphi_{\mathbf{l}} \qquad \Phi_4 = \psi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}}$$
$$\Phi_5 = \varphi_{\mathbf{h}} + \psi_{\mathbf{k}} + \psi_{\mathbf{l}} \qquad \Phi_6 = \psi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \psi_{\mathbf{l}}$$
$$\Phi_7 = \psi_{\mathbf{h}} + \psi_{\mathbf{k}} + \varphi_{\mathbf{l}} \qquad \Phi_8 = \psi_{\mathbf{h}} + \psi_{\mathbf{k}} + \psi_{\mathbf{l}}.$$

So, for the estimation of any Φ_j, the joint probability distribution

$$P(E_{\mathbf{h}}, E_{\mathbf{k}}, E_{\mathbf{l}}, G_{\mathbf{h}}, G_{\mathbf{k}}, G_{\mathbf{l}})$$

has to be studied, from which eight conditional probability densities can be obtained:

$$P(\Phi_i \big| |E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{l}}|, |G_{\mathbf{h}}|, |G_{\mathbf{k}}|, |G_{\mathbf{l}}|)$$
$$\simeq [2\pi I_0(Q_j)]^{-1} \exp[Q_j \cos \Phi_j]$$

for $j = 1, \ldots, 8$.

The analytical expressions of Q_j are too intricate and are not given here (the reader is referred to the original paper). We only say that Q_j may be positive or negative, so that reliable triplet phase estimates near 0 or near π are possible: the larger $|Q_j|$, the more reliable the phase estimate.

A useful interpretation of the formulae in terms of experimental parameters was suggested by Fortier *et al.* (1984): according to them, distributions do not depend, as in the case of the traditional three-phase invariants, on the total number of atoms per unit cell but rather on the scattering difference between the native protein and the derivative (that is, on the scattering of the heavy atoms in the derivative).

Hauptman's formulae were generalized by Giacovazzo *et al.* (1988): the new expressions were able to take into account the resolution effects on distribution parameters. The formulae are completely general and include as special cases native protein and heavy-atom isomorphous derivatives as well as X-ray and neutron diffraction data. Their complicated algebraic forms are easily reduced to a simple expression in the case of a native protein heavy-atom derivative: in particular, the reliability parameter for Φ_1 is

$$Q_1 = 2[\sigma_3/\sigma_2^{3/2}]_{\mathrm{p}} |E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{l}}| + 2[\sigma_3/\sigma_2^{3/2}]_{\mathrm{H}} \Delta_{\mathbf{h}} \Delta_{\mathbf{k}} \Delta_{\mathbf{l}}, \quad (2.2.10.2)$$

where indices p and H warn that parameters have to be calculated over protein atoms and over heavy atoms, respectively, and

$$\Delta = (F_{\mathrm{d}} - F_{\mathrm{p}})/(\textstyle\sum f_j^2)_{\mathrm{H}}^{1/2}.$$

Δ is a pseudo-normalized difference (with respect to the heavy-atom structure) between moduli of structure factors.

Equation (2.2.10.2) may be compared with Karle's (1983) algebraic rule: if the sign of $\Delta_{\mathbf{h}} \Delta_{\mathbf{k}} \Delta_{\mathbf{l}}$ is plus then the value of Φ_1 is estimated to be zero; if its sign is minus then the expected value of Φ_1 is close to π. In practice Karle's rule agrees with (2.2.10.2) only if the Cochran-type term in (2.2.10.2) may be neglected. Furthermore, (2.2.10.2) shows that large reliability values do not depend on the triple product of structure-factor differences, but on the triple product of pseudo-normalized differences.

A similar mathematical approach has been applied to estimate quartet invariants *via* isomorphous data. The result may be summarized as follows: a quartet is a phase relationship of order N_{H}^{-1} (Giacovazzo & Siliqi, 1996a,b; see also Kyriakidis *et al.*, 1996), with reliability factor equal to

$$G = \frac{2\Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{l}}\Delta_{\mathbf{h+k+l}}}{Q_4 N_{\mathrm{H}}}$$
$$\times \{1 + (\Delta_{\mathbf{h+k}}^2 - 1) + (\Delta_{\mathbf{h+l}}^2 - 1) + (\Delta_{\mathbf{k+l}}^2 - 1)\},$$

$$(2.2.10.3)$$

where Q_4 is a suitable normalizing factor.

As previously stressed, equations (2.2.10.2) and (2.2.10.3) are valid if the lack of isomorphism and the errors in the measurements are assumed to be negligible. At first sight this approach seems more appealing than the traditional two-step procedures, however it did not prove to be competitive with them. The main reason is the absence in the Hauptman and Giacovazzo approaches of a probabilistic treatment of the errors: such a treatment, on the contrary, is basic for the traditional SIR–MIR techniques [see Blow & Crick (1959) and Terwilliger & Eisenberg (1987) for two related approaches].

The problem of the errors in the probabilistic scenario defined by the joint probability distribution functions approach has recently been overcome by Giacovazzo *et al.* (2001). In their probabilistic calculations the following assumptions were made:

$$|F_{\mathrm{d}j}| \exp(i\varphi_j) = |F_{\mathrm{p}}| \exp(i\varphi_{\mathrm{p}}) + F_{\mathrm{H}j} \exp(i\varphi_{\mathrm{H}j}) + |\mu_j| \exp(i\theta_j),$$

$$(2.2.10.4)$$

where j refers to the jth derivative. $|\mu_j| \exp(i\theta_j)$ is the error, which can include model as well as measurement errors.

A more realistic expression for the reliability factor G of triplet invariants is obtained by including the expression (2.2.10.4) in the probabilistic approach. Then the reliability parameter of the triplet invariants is transformed into (Giacovazzo *et al.*, 2001)

$$G = 2[\sigma_3/\sigma_2^{3/2}]_{\mathrm{p}} R_{\mathrm{p}1} R_{\mathrm{p}2} R_{\mathrm{p}3}$$
$$+ 2[\sigma_3/\sigma_2^{3/2}]_{\mathrm{H}} \frac{\Delta_1 \Delta_2 \Delta_3}{[1 + (\sigma_{\mu 1}^2)_{\mathrm{H}}][1 + (\sigma_{\mu 2}^2)_{\mathrm{H}}][1 + (\sigma_{\mu 3}^2)_{\mathrm{H}}]},$$

$$(2.2.10.5)$$

where $(\sigma_\mu^2)_{\mathrm{H}} = |\mu|^2/(\Sigma f_j^2)_{\mathrm{H}}$.

Equation (2.2.10.5) suggests how the error influences the reliability of the triplet estimate: even quite a small value of $|\mu|^2$ may be critical if the scattering power of the heavy-atom substructure is a very small percentage of the derivative scattering power.

A one-step procedure has been implemented in a computer program (Giacovazzo *et al.*, 2002): it has been shown that the method is able to derive automatically, from the experimental data and without any user intervention, good quality (*i.e.* perfectly interpretable) electron-density maps.

2.2.10.5. SIR–MIR case: the two-step procedure. Finding the heavy-atom substructure by direct methods

The first trials for finding the heavy-atom substructure were based on the following assumption: the modulus of the isomorphous difference,

$$\Delta F = |F_{\mathrm{d}}| - |F_{\mathrm{p}}|,$$

is assumed at a first approximation as an estimate of the heavy-atom structure factor F_{H}. Perutz (1956) approximated $|F_{\mathrm{H}}|^2$ with the difference $(|F_{\mathrm{d}}|^2 - |F_{\mathrm{p}}|^2)$. Blow (1958) and Rossmann (1960) suggested a better approximation: $|F_{\mathrm{H}}|^2 \simeq |\Delta F|^2$. A deeper analysis was performed by Phillips (1966), Dodson & Vijayan (1971), Blessing & Smith (1999) and Grosse-Kunstleve & Brunger (1999). The use of direct methods requires the normalization of $|\Delta F|$ and application of the tangent formula (Wilson, 1978).

A sounder procedure has been suggested by Giacovazzo *et al.* (2004): they studied, for the SIR case, the joint probability distribution function

$$P(E_{\mathrm{H}}, E_{\mathrm{p}}, E_{\mathrm{d}})$$

under the following assumptions:

(*a*) the atomic positions of the native protein structure and the positions of the heavy atoms in the derivative structure are the primitive random variables of the probabilistic approach;

(*b*)

$$|F_{\mathrm{d}}| \exp(i\varphi_{\mathrm{d}}) = |F_{\mathrm{p}}| \exp(i\varphi_{\mathrm{p}}) + |F_{\mathrm{H}}| \exp(i\varphi_{\mathrm{H}}) + |\mu_{\mathrm{d}}| \exp(i\theta_{\mathrm{d}})$$

$$(2.2.10.6)$$

is the structure factor of the derivative.

Then the conditional distribution $P(R_{\mathrm{H}}|R_{\mathrm{p}}, R_{\mathrm{d}})$ may be derived, from which $\langle R_{\mathrm{H}}|R_{\mathrm{p}}, R_{\mathrm{d}}\rangle$ may be obtained. In terms of structure factors

$$\langle|F_{\mathrm{H}}|^2\rangle = \frac{\sum_{\mathrm{H}}}{(\sum_{\mathrm{H}} + \langle|\mu_{\mathrm{d}}|^2\rangle)} \left[\langle|\mu_{\mathrm{d}}|^2\rangle + \frac{\sum_{\mathrm{H}}}{(\sum_{\mathrm{H}} + \langle|\mu_{\mathrm{d}}|^2\rangle)} \Delta_{\mathrm{iso}}^2 \right].$$

$$(2.2.10.7)$$

The effect of the errors on the evaluation of the moduli $|F_{\mathrm{H}}|^2$ may be easily derived: if $\langle|\mu_{\mathrm{d}}|^2\rangle = 0$, equation (2.2.10.7) confirms Blow and Rossmann's approximation $\langle|F_{\mathrm{H}}|^2\rangle \simeq |\Delta F|^2$. If $\langle|\mu_{\mathrm{d}}|^2\rangle \neq 0$ Blow and Rossmann's estimate should be affected by a systematic error, increasing with $\langle|\mu_{\mathrm{d}}|^2\rangle$.

2.2.10.6. SIR–MIR case: protein phasing by direct methods

Let us suppose that the various heavy-atom substructures have been determined. They may be used as additional prior information for a more accurate estimate of the φ_{p} values. To this purpose the distributions

$$P(E_{\mathrm{p}}, \mathbf{E}_{\mathrm{d}}'|\mathbf{E}_{\mathrm{H}}') \equiv P(E_{\mathrm{p}}, E_{\mathrm{d}1}', \dots, E_{\mathrm{d}n}'|E_{\mathrm{H}1}', \dots, E_{\mathrm{H}n}') \quad (2.2.10.8)$$

may be used under the assumption (2.2.10.6). $E_{\mathrm{d}j}'$ and $E_{\mathrm{H}j}'$, for $j = 1, \dots, n$, are the structure factors of the jth derivative and of the jth heavy-atom substructure, respectively, both normalized with respect to the protein. Any joint probability density (2.2.10.8) may be reliably approximated by a multidimensional Gaussian distribution (Giacovazzo & Siliqi, 2002), from which the following conditional distribution is obtained:

$$P(\varphi_{\mathrm{p}}|R_{\mathrm{p}}, \mathbf{R}_{\mathrm{d}}', \mathbf{E}_{\mathrm{H}}') \simeq [2\pi I_0(G)]^{-1} \exp[\alpha_{\mathrm{p}} \cos(\varphi_{\mathrm{p}} - \theta_{\mathrm{p}})]$$

where θ_{p}, the expected value of φ_{p}, is given by

$$\tan \theta_{\mathrm{p}} = \frac{\sum_{j=1}^n G_j \sin \varphi_{\mathrm{H}j}}{\sum_{j=1}^n G_j \cos \varphi_{\mathrm{H}j}} = \frac{T}{B}$$

and $G_j = 2|F_{\mathrm{H}j}|\Delta F/\mu_j^2$.

$\alpha_{\mathrm{p}} = (T^2 + B^2)^{1/2}$ is the reliability factor of the phase estimate. A robust phasing procedure has been established which, starting from the observed moduli $|F_{\mathrm{p}}|, |F_{\mathrm{d}j}|, j = 1, \dots, n$, is able to automatically provide, without any user intervention, a high-quality electron-density map of the protein (Giacovazzo *et al.*, 2002).

2.2.10.7. Integration of anomalous-dispersion techniques with direct methods

If the frequency of the radiation is close to an absorption edge of an atom, then that atom will scatter the X-rays anomalously (see Chapter 2.4) according to $f = f' + if''$. This results in the breakdown of Friedel's law. It was soon realized that the Bijvoet difference could also be used in the determination of phases (Peerdeman & Bijvoet, 1956; Ramachandran & Raman, 1956;

Okaya & Pepinsky, 1956). Since then, a great deal of work has been done both from algebraic (see Chapter 2.4) and from probabilistic points of view. In this section we are only interested in the second.

SAD (single anomalous dispersion) and MAD (multiple anomalous dispersion) techniques can be used. Both are characterized by one protein structure and one anomalous-scatterer substructure. The experimental diffraction data differ only because of the different anomalous scattering (not because of different anomalous-scatterer substructures). In the MAD case the anomalous-scatterer substructure is in some way 'over-determined' by the data and, therefore, it is more convenient to use a two-step procedure: first define the positions of the anomalous scatterers, and then estimate the protein phase values. For completeness, we describe the one-step procedures in Section 2.2.10.8. These are based on the estimation of the structure invariants and on the application of the tangent formula. The two-step procedures are described in the Sections 2.2.10.9 and 2.2.10.10.

2.2.10.8. The SAD case: the one-step procedures

Probability distributions of diffraction intensities and of selected functions of diffraction intensities for dispersive structures have been given by various authors [Parthasarathy & Srinivasan (1964), see also Srinivasan & Parthasarathy (1976) and relevant literature cited therein]. We describe here some probabilistic formulae for estimating invariants of low order.

(*a*) *Estimation of two-phase structure invariants.* The conditional probability distribution of $\Phi = \varphi_{\mathbf{h}} + \varphi_{-\mathbf{h}}$ given $R_{\mathbf{h}}$ and $G_{\mathbf{h}}$ (normalized moduli of $F_{\mathbf{h}}$ and $F_{-\mathbf{h}}$, respectively) (Hauptman, 1982*b*; Giacovazzo, 1983*b*) is

$$P(\Phi | R_{\mathbf{h}}, G_{\mathbf{h}}) \simeq [2\pi I_0(Q)]^{-1} \exp[Q \cos(\Phi - q)], \quad (2.2.10.9)$$

where

$$Q = \frac{2R_{\mathbf{h}} G_{\mathbf{h}}}{\sqrt{c}} [c_1^2 + c_2^2]^{1/2},$$

$$\cos q = \frac{c_1}{[c_1^2 + c_2^2]^{1/2}}, \qquad \sin q = \frac{c_2}{[c_1^2 + c_2^2]^{1/2}},$$

$$c_1 = \sum_{j=1}^{N} (f_j'^2 - f_j''^2) / \sum,$$

$$c_2 = 2 \sum_{j=1}^{N} f_j' f_j'' / \sum,$$

$$c = [1 - (c_1^2 + c_2^2)]^2,$$

$$\sum = \sum_{j=1}^{N} (f_j'^2 + f_j''^2).$$

q is the most probable value of Φ: a large value of the parameter Q suggests that the phase relation $\Phi = q$ is reliable. Large values of Q are often available in practice: q, however, may be considered an estimate of $|\Phi|$ rather than of Φ because the enantiomorph is not fixed in (2.2.10.9). A formula for the estimation of Φ in centrosymmetric structures has been provided by Giacovazzo (1987).

(*b*) *Estimation of triplet invariants.* Kroon *et al.* (1977) first incorporated anomalous diffraction in order to estimate triplet invariants. Their work was based on an analysis of the complex double Patterson function. Subsequent probabilistic considerations (Heinermann *et al.*, 1978) confirmed their results, which can be so expressed:

$$\sin \bar{\Phi} = \frac{|\tau|^2 - |\bar{\tau}|^2}{4\tau''[\frac{1}{2}(|\tau|^2 + |\bar{\tau}|^2) - |\tau''|^2]^{1/2}}, \quad (2.2.10.10)$$

where $(\mathbf{h} + \mathbf{k} + \mathbf{l} = 0)$,

$$\tau = E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{l}} = R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{l}} \exp(i\Phi_{\mathbf{h}, \mathbf{k}}),$$

$$\bar{\tau} = E_{-\mathbf{h}} E_{-\mathbf{k}} E_{-\mathbf{l}} = G_{\mathbf{h}} G_{\mathbf{k}} G_{\mathbf{l}} \exp(i\Phi_{\bar{\mathbf{h}}, \bar{\mathbf{k}}}),$$

$$\bar{\Phi} = \tfrac{1}{2}(\Phi_{\mathbf{h}, \mathbf{k}} - \Phi_{\bar{\mathbf{h}}, \bar{\mathbf{k}}}),$$

and τ'' is the contribution of the imaginary part of τ, which may be approximated in favourable conditions by

$$\tau'' = 2f''[f_{\mathbf{h}}' f_{\mathbf{k}}' + f_{\mathbf{h}}' f_{\mathbf{l}}' + f_{\mathbf{k}}' f_{\mathbf{l}}] \\ \times [1 + S(R_{\mathbf{h}}^2 + R_{\mathbf{k}}^2 + R_{\mathbf{l}}^2 - 3)],$$

where S is a suitable scale factor.

Equation (2.2.10.10) gives two possible values for $\bar{\Phi}$ (Φ and $\pi - \Phi$). Only if $R_{\mathbf{h}} R_{\mathbf{k}} R_{\overline{\mathbf{h+k}}}$ is large enough may this phase ambiguity be resolved by choosing the angle nearest to zero.

The evaluation of triplet phases by means of anomalous dispersion has been further pursued by Hauptman (1982*b*) and independently by Giacovazzo (1983*b*). Owing to the breakdown of Friedel's law there are eight distinct triplet invariants which can contemporaneously be exploited:

$$\Phi_1 = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}}, \qquad \Phi_2 = -\varphi_{-\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}}$$

$$\Phi_3 = \varphi_{\mathbf{h}} - \varphi_{-\mathbf{k}} + \varphi_{\mathbf{l}}, \qquad \Phi_4 = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{-\mathbf{l}}$$

$$\Phi_5 = \varphi_{-\mathbf{h}} + \varphi_{-\mathbf{k}} + \varphi_{-\mathbf{l}}, \qquad \Phi_6 = -\varphi_{\mathbf{h}} + \varphi_{-\mathbf{k}} + \varphi_{-\mathbf{l}}$$

$$\Phi_7 = \varphi_{-\mathbf{h}} - \varphi_{\mathbf{k}} + \varphi_{-\mathbf{l}}, \qquad \Phi_8 = \varphi_{-\mathbf{h}} + \varphi_{-\mathbf{k}} - \varphi_{\mathbf{l}}.$$

Given

$$R_1 = |E_{\mathbf{h}}| \qquad R_2 = |E_{\mathbf{k}}| \qquad R_3 = |E_{\mathbf{h+k}}|$$

$$G_1 = |E_{-\mathbf{h}}| \qquad G_2 = |E_{-\mathbf{k}}| \qquad R_3 = |E_{-\mathbf{h-k}}|$$

$$\varphi_1 = \varphi_{\mathbf{h}} \qquad \varphi_2 = \varphi_{\mathbf{k}} \qquad \varphi_3 = \varphi_{\mathbf{h+k}}$$

$$\psi_1 = \varphi_{-\mathbf{h}} \qquad \psi_2 = \varphi_{-\mathbf{k}} \qquad \psi_3 = \varphi_{-\mathbf{h-k}},$$

Hauptman and Giacovazzo found the following conditional distribution:

$$P(\Phi | R_j, G_j, j = 1, 2, 3) \simeq \left[2\pi I_0(\Omega)\right]^{-1} \exp[\Omega \cos(\Phi - \omega)]. \quad (2.2.10.11)$$

The definitions of Ω and ω are rather extensive and so the reader is referred to the published papers. We only add that Ω is always positive and that ω, the expected value of Φ, may lie anywhere between 0 and 2π. Understanding the role of the various parameters in equation (2.2.10.11) is not easy. Giacovazzo *et al.* (2003) found an equivalent simpler expression from which interpretable estimates of the parameters were obtained. In the same paper the limitations of the approach (*versus* the two-step procedures) were clarified.

2.2.10.9. SAD–MAD case: the two-step procedures. Finding the anomalous-scatterer substructure by direct methods

The anomalous-scatterer substructure is traditionally determined by the techniques suggested by Karle and Hendrickson (Karle, 1980*b*; Hendrickson, 1985; Pähler *et al.*, 1990; Terwilliger, 1994). The introduction of selenium into proteins as seleno-methionine encouraged the second-generation direct methods programs [*Shake and Bake* by Miller *et al.* (1994); *Half bake* by Sheldrick (1998); *SIR2000-N* by Burla *et al.* (2001); *ACORN* by Foadi *et al.* (2000)] to locate Se atoms. Since the number of Se atoms may be quite large (up to 200), direct methods rather than Patterson techniques seem to be preferable. *Shake and Bake*, *Half Bake* and *ACORN* obtain the coordinates of the anomalous scatterers from a single-wavelength set of data. When more sets of diffraction data are available the solutions obtained by the other sets are used to confirm the correct solution.

A different approach has been suggested in two recent papers (Burla *et al.*, 2002; Burla, Carrozzini *et al.*, 2003): the estimates of the amplitudes of the structure factors of the anomalously scattering substructure are derived, *via* the rigorous method of the joint probability distribution functions, from the experimental diffraction moduli relative to n wavelengths. To do that, first the joint distribution

$$P_n = P(A_{oa}, A_1^+, A_2^+, \ldots, A_n^+, A_1^-, A_2^-, \ldots, A_n^-,$$
$$B_{oa}, B_1^+, B_2^+, \ldots, B_n^+, B_1^-, B_2^-, \ldots, B_n^-)$$
$$= \pi^{-(2n+1)}(\det K)^{1/2} \exp(-\tfrac{1}{2}\mathbf{T}^T \mathbf{K}^{-1} \mathbf{T})$$

is calculated, where $A_{oa}, B_{oa}, E_{oa}, A_i^+, B_i^+, A_i^-, B_i^-$ are the real and imaginary components of E_{oa}, E_i^+, E_i^-, respectively, $\mathbf{K}$ is a symmetric square matrix of order $(4n + 2)$, $\mathbf{K}^{-1} = \{\lambda_{ij}\}$ is its inverse, and $\mathbf{T}$ is a suitable vector with components defined in terms of the variables $A_{oa}, A_1^+, A_2^+, \ldots, B_n^-$. E_{oa} is the normalized structure factor of the anomalous scatterer substructure calculated by neglecting anomalous scattering components. Then the conditional distribution

$$P(R_{oa} | R_1, \ldots, R_n, G_1, \ldots, G_n)$$

is derived, from which

$$\langle R_{oa} | R_1, \ldots, G_n \rangle = \tfrac{1}{2}(\pi/\lambda_{11})^{1/2}[1 + 4X^2/(\pi\lambda_{11})]^{1/2} \quad (2.2.10.12)$$

is obtained, where

$$X^2 = Q_1^2 + Q_2^2$$
$$Q_1 = \lambda_{12}R_1 + \lambda_{13}R_2 + \ldots + \lambda_{1,n+1}R_n + \lambda_{1,n+2}G_1 + \ldots$$
$$\quad + \lambda_{1,2n+1}G_n$$
$$Q_2 = \lambda_{1,2n+3}R_1 + \lambda_{1,2n+4}R_2 + \ldots + \lambda_{1,3n+2}R_n + \ldots - \lambda_{1,3n+3}G_1$$
$$\quad - \ldots - \lambda_{1,4n+2}G_n.$$

The standard deviation of the estimate is also calculated:

$$\sigma_{R_{oa}} = \left[\langle R_{oa}^2 | \ldots \rangle - \langle R_{oa} | \ldots \rangle^2 \right]^{1/2} = \left[\left(1 - \frac{\pi}{4}\right)\lambda_{11}^{-1} \right]^{1/2},$$

from which

$$\frac{\langle R_{oa} | \ldots \rangle}{\sigma_{R_{oa}}} = \left[\frac{(\pi/4) + (X^2)/\lambda_{11}}{1 - (\pi/4)} \right]^{1/2}. \quad (2.2.10.13)$$

The advantage of the above approach is that the estimates can simultaneously exploit both the anomalous and the dispersive differences. The computing procedure proposed by Burla, Carrozzini *et al.* (2003) is the following:

(i) The sets S_j, $j = 1, \ldots, n$, of the observed magnitudes (say $|F^+|, |F^-|$) are stored for all the n wavelengths.

(ii) The Wilson method is applied to put the sets S_j on their absolute scales.

(iii) Equations (2.2.10.12) and (2.2.10.13) are applied to obtain the values $\langle R_{oa} | \ldots \rangle$ and $\langle R_{oa} | \ldots \rangle / \sigma_{R_{oa}}$.

(iv) The triplet invariants involving the reflections with the highest $\langle R_{oa} | \ldots \rangle / \sigma_{R_{oa}}$ values are evaluated and the tangent formula is applied *via* a random starting approach.

(v) The direct-space refinement techniques of *SIR2002* (Burla, Camalli *et al.*, 2003) are used to extend the phase information to a larger set of reflections: only 30% of the reflections with the smallest values of $\langle R_{oa} | \ldots \rangle$ remain unphased. Automatic cycles of least-squares refinement improve the substructure model provided by the trial solutions.

(vi) Suitable figures of merit are used to recognize the correct substructure models.

The application of the above procedure to several MAD cases showed that the various wavelength combinations are not equally informative. A criterion based on the correlation among the various Δ_{ano} values was also provided (see also Schneider & Sheldrick, 2002) for predicting the most informative combinations.

2.2.10.10. SAD–MAD case: protein phasing by direct methods

Once the anomalous-scatterer substructure has been found, the corresponding structure factors $E_{a1}^+, \ldots, E_{an}^+, E_{a1}^-, \ldots, E_{an}^-$ are known in modulus and phase. Then the conditional joint probability distribution

$$P\big(E_1^+, \ldots, E_n^+, E_1^-, \ldots, E_n^- | E_{a1}^+, \ldots, E_{an}^+, E_{a1}^-, \ldots, E_{an}^-\big)$$

may be calculated (Giacovazzo & Siliqi, 2004), from which the conditional distribution

$$P\big(\varphi_1^+ | E_{ai}^+, E_{ai}^-, R_i, G_i, i = 1, \ldots, 2\big)$$

may be derived.

It has been shown that the most probable phase of φ_1^+, say θ_1^+, is the phase of the vector

$$\sum_{j=1}^{n} \left[w_j^+ E_{aj}^+ + w_j^- E_{aj}^{-*} \right]$$
$$+ \sum_{j,p=1, p>j}^{n} \left[w_{jp}(E_{aj}^+ - E_{ap}^+) + w_{n+j,n+p}\big(E_{aj}^{-*} - E_{ap}^{-*}\big) \right]$$
$$+ \sum_{j,p=1}^{n} w_{j,n+p}\big(E_{aj}^+ - E_{ap}^{-*}\big)] \quad (2.2.10.14)$$

and the reliability parameter of the phase estimate is nothing other than the modulus of (2.2.10.14). The first term in (2.2.10.14) is a Sim-like contribution; the other terms, through the weights w, take into account the errors and the experimental differences $(R_j - R_p)$, $(G_j - G_p)$ and $(R_j - G_p)$.

References

Allegra, G. (1979). *Derivation of three-phase invariants from the Patterson function. Acta Cryst.* A**35**, 213–220.

Altomare, A., Burla, M. C., Camalli, M., Cascarano, G. L., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Polidori, G. & Spagna, R. (1999). *SIR97: a new tool for crystal structure determination and refinement. J. Appl. Cryst.* **32**, 115–119.

Anzenhofer, K. & Hoppe, W. (1962). *Phys. Verh. Mosbach.* **13**, 119.

Ardito, G., Cascarano, G., Giacovazzo, C. & Luić, M. (1985). *1-Phase seminvariants and Harker sections. Z. Kristallogr.* **172**, 25–34.

Argos, P. & Rossmann, M. G. (1980). *Molecular replacement method.* In *Theory and Practice of Direct Methods in Crystallography,* edited by M. F. C. Ladd & R. A. Palmer, pp. 381–389. New York: Plenum.

Avrami, M. (1938). *Direct determination of crystal structure from X-ray data. Phys. Rev.* **54**, 300–303.

Baggio, R., Woolfson, M. M., Declercq, J.-P. & Germain, G. (1978). *On the application of phase relationships to complex structures. XVI. A random approach to structure determination. Acta Cryst.* A**34**, 883–892.

Banerjee, K. (1933). *Determination of the signs of the Fourier terms in complete crystal structure analysis. Proc. R. Soc. London Ser. A,* **141**, 188–193.

Bertaut, E. F. (1955a). *La méthode statistique en cristallographie. I. Acta Cryst.* **8**, 537–543.

Bertaut, E. F. (1955b). *La méthode statistique en cristallographie. II. Quelques applications. Acta Cryst.* **8**, 544–548.

Bertaut, E. F. (1960). *Ordre logarithmique des densités de répartition. I. Acta Cryst.* **13**, 546–552.

Beurskens, P. T., Beurskens, G., de Gelder, R., Garcia-Granda, S., Gould, R. O., Israel, R. & Smits, J. M. M. (1999). *The DIRDIF-99 program system.* Crystallography Laboratory, University of Nijmegen, The Netherlands.

Beurskens, P. T., Gould, R. O., Bruins Slot, H. J. & Bosman, W. P. (1987). *Translation functions for the positioning of a well oriented molecular fragment. Z. Kristallogr.* **179**, 127–159.

Beurskens, P. T., Prick, A. J., Doesburg, H. M. & Gould, R. O. (1979). *Statistical properties of normalized difference-structure factors for non-centrosymmetric structures. Acta Cryst.* A**35**, 765–772.

Bhuiya, A. K. & Stanley, E. (1963). *The refinement of atomic parameters by direct calculation of the minimum residual. Acta Cryst.* **16**, 981–984.

Blessing, R. H. & Smith, G. D. (1999). *Difference structure-factor normalization for heavy-atom or anomalous-scattering substructure determinations. J. Appl. Cryst.* **32**, 664–670.

Blow, D. M. (1958). *An X-ray examination of some crystal forms of pig and rabbit haemoglobin. Acta Cryst.* **11**, 125–126.

Blow, D. M. & Crick, F. H. C. (1959). *The treatment of errors in the isomorphous replacement method. Acta Cryst.* **12**, 794–802.

Böhme, R. (1982). *Direkte Methoden für Strukturen mit Uberstruktur-effekten. Acta Cryst.* A**38**, 318–326.

Bouman, J. (1956). *A general theory of inequalities. Acta Cryst.* **9**, 777–780.

Bricogne, G. (1984). *Maximum entropy and the foundation of direct methods. Acta Cryst.* A**40**, 410–415.

Britten, P. L. & Collins, D. M. (1982). *Information theory as a basis for the maximum determinant. Acta Cryst.* A**38**, 129–132.

Buerger, M. J. (1959). *Vector Space and its Applications in Crystal Structure Investigation.* New York: John Wiley.

Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Polidori, G. & Spagna, R. (2005). *SIR2004: an improved tool for crystal structure determination and refinement. J. Appl. Cryst.* **38**, 381–388.

Burla, M. C., Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Polidori, G. (2004). *Ab initio protein phasing: the Patterson deconvolution method in SIR2002. J. Appl. Cryst.* **37**, 258–264.

Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (2001). *SIR2000-N, a program for large and small crystal structures. J. Appl. Cryst.* **34**, 523–526.

Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (2003). *SIR2002: the program. J. Appl. Cryst.* **36**, 1103.

Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2003). *SAD or MAD phasing: location of the anomalous scatterers. Acta Cryst.* D**59**, 662–669.

Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Siliqi, D. (2002). *MAD phasing: probabilistic estimate of |F_{oa}|. Acta Cryst.* D**58**, 928–935.

Burla, M. C., Cascarano, G., Giacovazzo, C., Nunzi, A. & Polidori, G. (1987). *A weighting scheme for tangent formula development. III. The weighting scheme of the SIR program. Acta Cryst.* A**43**, 370–374.

Busetta, B. (1976). *An example of the use of quartet and triplet structure invariants when enantiomorph discrimination is difficult. Acta Cryst.* A**32**, 139–143.

Busetta, B., Giacovazzo, C., Burla, M. C., Nunzi, A., Polidori, G. & Viterbo, D. (1980). *The SIR program. I. Use of negative quartets. Acta Cryst.* A**36**, 68–74.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005a). *Phasing at resolution higher than the experimental resolution. Acta Cryst.* D**61**, 556–565.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005b). *Ab initio phasing at resolution higher than experimental resolution. Acta Cryst.* D**61**, 1080–1087.

Camalli, M., Giacovazzo, C. & Spagna, R. (1985). *From a partial to the complete crystal structure. II. The procedure and its applications. Acta Cryst.* A**41**, 605–613.

Cascarano, G. & Giacovazzo, C. (1983). *One-phase seminvariants of first rank. I. Algebraic considerations. Z. Kristallogr.* **165**, 169–174.

Cascarano, G., Giacovazzo, C., Burla, M. C., Nunzi, A. & Polidori, G. (1984). *The distribution of α_h. Acta Cryst.* A**40**, 389–394.

Cascarano, G., Giacovazzo, C., Calabrese, G., Burla, M. C., Nunzi, A., Polidori, G. & Viterbo, D. (1984). *One-phase seminvariants of first rank. II. Probabilistic considerations. Z. Kristallogr.* **167**, 37–47.

Cascarano, G., Giacovazzo, C., Camalli, M., Spagna, R., Burla, M. C., Nunzi, A. & Polidori, G. (1984). *The method of representations of structure seminvariants. The strengthening of triplet relationships. Acta Cryst.* A**40**, 278–283.

Cascarano, G., Giacovazzo, C. & Luić, M. (1985a). *Non-crystallographic translational symmetry: effects on diffraction-intensity statistics.* In *Structure and Statistics in Crystallography*, edited by A. J. C. Wilson, pp. 67–77. Guilderland, USA: Adenine Press.

Cascarano, G., Giacovazzo, C. & Luić, M. (1985b). *Direct methods and superstructures. I. Effects of the pseudotranslation on the reciprocal space. Acta Cryst.* A**41**, 544–551.

Cascarano, G., Giacovazzo, C. & Luić, M. (1987). *Direct methods and structures showing superstructure effects. II. A probabilistic theory of triplet invariants. Acta Cryst.* A**43**, 14–22.

Cascarano, G., Giacovazzo, C. & Luić, M. (1988a). *Direct methods and structures showing superstructure effects. III. A general mathematical model. Acta Cryst.* A**44**, 176–183.

Cascarano, G., Giacovazzo, C. & Luić, M. (1988b). *Direct methods and structures showing superstructure effects. IV. A new approach for phase solution. Acta Cryst.* A**44**, 183–188.

Cascarano, G., Giacovazzo, C., Luić, M., Pifferi, A. & Spagna, R. (1987). *1-Phase seminvariants and Harker sections. II. A new procedure. Z. Kristallogr.* **179**, 113–125.

Cascarano, G., Giacovazzo, C. & Viterbo, D. (1987). *Figures of merit in direct methods: a new point of view. Acta Cryst.* A**43**, 22–29.

Castellano, E. E., Podjarny, A. D. & Navaza, J. (1973). *A multivariate joint probability distribution of phase determination. Acta Cryst.* A**29**, 609–615.

Cochran, W. (1955). *Relations between the phases of structure factors. Acta Cryst.* **8**, 473–478.

Cochran, W. & Douglas, A. S. (1957). *The use of a high-speed digital computer for the direct determination of crystal structure. II. Proc. R. Soc. London Ser. A*, **243**, 281–288.

Cochran, W. & Woolfson, M. M. (1955). *The theory of sign relations between structure factors. Acta Cryst.* **8**, 1–12.

Coulter, C. L. & Dewar, R. B. K. (1971). *Tangent formula applications in protein crystallography: an evaluation. Acta Cryst.* B**27**, 1730–1740.

Crowther, R. A. & Blow, D. M. (1967). *A method of positioning a known molecule in an unknown crystal structure. Acta Cryst.* **23**, 544–548.

Cutfield, J. F., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., Isaacs, N. W., Sakabe, K. & Sakabe, N. (1975). *The high resolution structure of insulin: a comparison of results obtained from least-squares phase refinement and difference Fourier refinement. Acta Cryst.* A**31**, S21.

Debaerdemaeker, T., Tate, C. & Woolfson, M. M. (1985). *On the application of phase relationships to complex structures. XXIV. The Sayre tangent formula. Acta Cryst.* A**41**, 286–290.

Debaerdemaeker, T. & Woolfson, M. M. (1972). *On the application of phase relationships to complex structures. IV. The coincidence method applied to general phases. Acta Cryst.* A**28**, 477–481.

Debaerdemaeker, T. & Woolfson, M. M. (1983). *On the application of phase relationships to complex structures. XXII. Techniques for random refinement. Acta Cryst.* A**39**, 193–196.

Declercq, J.-P., Germain, G., Main, P. & Woolfson, M. M. (1973). *On the application of phase relationships to complex structures. V. Finding the solution. Acta Cryst.* A**29**, 231–234.

Declercq, J.-P., Germain, G. & Woolfson, M. M. (1975). *On the application of phase relationships to complex structures. VIII. Extension of the magic-integer approach. Acta Cryst.* A**31**, 367–372.

De Titta, G. T., Edmonds, J. W., Langs, D. A. & Hauptman, H. (1975). *Use of the negative quartet cosine invariants as a phasing figure of merit: NQEST. Acta Cryst.* A**31**, 472–479.

DeTitta, G. T., Weeks, C. M., Thuman, P., Miller, R. & Hauptman, H. A. (1994). *Structure solution by minimal-function phase refinement and Fourier filtering. I. Theoretical basis. Acta Cryst.* A**50**, 203–210.

Dodson, E. & Vijayan, M. (1971). *The determination and refinement of heavy-atom parameters in protein heavy-atom derivatives. Some model calculations using acentric reflexions. Acta Cryst.* B**27**, 2402–2411.

Egert, E. (1983). *Patterson search – an alternative to direct methods. Acta Cryst.* A**39**, 936–940.

Egert, E. & Sheldrick, G. M. (1985). *Search for a fragment of known geometry by integrated Patterson and direct methods. Acta Cryst.* A**41**, 262–268.

Eller, G. von (1973). *Génération de formules statistiques entre facteurs de structure: la méthode du polynome. Acta Cryst.* A**29**, 63–67.

Fan, H.-F. (1999). *Crystallographic software: teXsan for Windows.* http://www.rigaku.com/downloads/journal/Vol15.1.1998/texsan.pdf.

Fan, H. F., Hao, Q. & Woolfson, M. M. (1991). *Proteins and direct methods. Z. Kristallogr.* **197**, 197–208.

Fan, H.-F., Yao, J.-X., Main, P. & Woolfson, M. M. (1983). *On the application of phase relationships to complex structures. XXIII. Automatic determination of crystal structures having pseudo-translational symmetry by a modified MULTAN procedure. Acta Cryst.* A**39**, 566–569.

Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *A flexible and efficient procedure for the solution and phase refinement of protein structures. Acta Cryst.* D**56**, 1137–1147.

Fortier, S. & Hauptman, H. (1977). *Quintets in P$\bar{1}$: probabilistic theory of the five-phase structure invariant in the space group P$\bar{1}$. Acta Cryst.* A**33**, 829–833.

Fortier, S., Weeks, C. M. & Hauptman, H. (1984). *On integrating the techniques of direct methods and isomorphous replacement. III. The three-phase invariant for the native and two-derivative case. Acta Cryst.* A**40**, 646–651.

Freer, A. A. & Gilmore, C. J. (1980). *The use of higher invariants in MULTAN. Acta Cryst.* A**36**, 470–475.

French, S. & Wilson, K. (1978). *On the treatment of negative intensity observations. Acta Cryst.* A**34**, 517–525.

Gelder, R. de (1992). Thesis. University of Leiden, The Netherlands.

Gelder, R. de, de Graaff, R. A. G. & Schenk, H. (1990). *On the construction of Karle–Hauptman matrices. Acta Cryst.* A**46**, 688–692.

Gelder, R. de, de Graaff, R. A. G. & Schenk, H. (1993). *Automatic determination of crystal structures using Karle–Hauptman matrices. Acta Cryst.* A**49**, 287–293.

Germain, G., Main, P. & Woolfson, M. M. (1970). *On the application of phase relationships to complex structures. II. Getting a good start. Acta Cryst.* B**26**, 274–285.

Germain, G., Main, P. & Woolfson, M. M. (1971). *The application of phase relationships to complex structures. III. The optimum use of phase relationships. Acta Cryst.* A**27**, 368–376.

Giacovazzo, C. (1974). *A new scheme for seminvariant tables in all space groups. Acta Cryst.* A**30**, 390–395.

Giacovazzo, C. (1975). *A probabilistic theory in P$\bar{1}$ of the invariant $E_\mathbf{h}E_\mathbf{k}E_\mathbf{l}E_\mathbf{h+k+l}$. Acta Cryst.* A**31**, 252–259.

Giacovazzo, C. (1976). *A probabilistic theory of the cosine invariant $\cos(\varphi_\mathbf{h} + \varphi_\mathbf{k} + \varphi_\mathbf{l} - \varphi_\mathbf{h+k+l})$. Acta Cryst.* A**32**, 91–99.

Giacovazzo, C. (1977a). *A general approach to phase relationships: the method of representations. Acta Cryst.* A**33**, 933–944.

Giacovazzo, C. (1977b). *Strengthening of the triplet relationships. II. A new probabilistic approach in P$\bar{1}$. Acta Cryst.* A**33**, 527–531.

Giacovazzo, C. (1977c). *On different probabilistic approaches to quartet theory. Acta Cryst.* A**33**, 50–54.

Giacovazzo, C. (1977d). *Quintets in P$\bar{1}$ and related phase relationships: a probabilistic approach. Acta Cryst.* A**33**, 944–948.

Giacovazzo, C. (1977e). *A probabilistic theory of the coincidence method. I. Centrosymmetric space groups. Acta Cryst.* A**33**, 531–538.

Giacovazzo, C. (1977f). *A probabilistic theory of the coincidence method. II. Non-centrosymmetric space groups. Acta Cryst.* A**33**, 539–547.

Giacovazzo, C. (1978). *The estimation of the one-phase structure seminvariants of first rank by means of their first and second representation. Acta Cryst.* A**34**, 562–574.

Giacovazzo, C. (1979a). *A probabilistic theory of two-phase seminvariants of first rank via the method of representations. III. Acta Cryst.* A**35**, 296–305.

Giacovazzo, C. (1979b). *A theoretical weighting scheme for tangent-formula development and refinement and Fourier synthesis. Acta Cryst.* A**35**, 757–764.

Giacovazzo, C. (1980a). *Direct Methods in Crystallography.* London: Academic Press.

Giacovazzo, C. (1980b). *The method of representations of structure seminvariants. II. New theoretical and practical aspects. Acta Cryst.* A**36**, 362–372.

Giacovazzo, C. (1980c). *Triplet and quartet relations: their use in direct procedures. Acta Cryst.* A**36**, 74–82.

Giacovazzo, C. (1983a). *From a partial to the complete crystal structure. Acta Cryst.* A**39**, 685–692.

Giacovazzo, C. (1983b). *The estimation of two-phase invariants in P$\bar{1}$ when anomalous scatterers are present. Acta Cryst.* A**39**, 585–592.

Giacovazzo, C. (1987). *One wavelength technique: estimation of centrosymmetrical two-phase invariants in dispersive structures. Acta Cryst.* A**43**, 73–75.

Giacovazzo, C. (1988). *New probabilistic formulas for finding the positions of correctly oriented atomic groups. Acta Cryst.* A**44**, 294–300.

Giacovazzo, C. (1998). *Direct Phasing in Crystallography.* New York: IUCr, Oxford University Press.

Giacovazzo, C., Cascarano, G. & Zheng, C.-D. (1988). *On integrating the techniques of direct methods and isomorphous replacement. A new probabilistic formula for triplet invariants. Acta Cryst.* A**44**, 45–51.

Giacovazzo, C., Guagliardi, A., Ravelli, R. & Siliqi, D. (1994). *Ab initio direct phasing of proteins: the limits. Z. Kristallogr.* **209**, 136–142.

Giacovazzo, C., Ladisa, M. & Siliqi, D. (2002). *Crystal structure solution of proteins by direct methods: an automatic procedure for SIR–MIR and SIRAS–MIRAS cases. Acta Cryst.* A**58**, 598–604.

Giacovazzo, C., Ladisa, M. & Siliqi, D. (2003). *The estimation of three-phase invariants when anomalous scatterers are present: the limits. Acta Cryst.* A**59**, 569–576.

Giacovazzo, C., Moustiakimov, M., Siliqi, D. & Pifferi, A. (2004). *Locating heavy atoms by integrating direct methods and SIR techniques. Acta Cryst.* A**60**, 233–238.

Giacovazzo, C. & Siliqi, D. (1996a). *On integrating direct methods and isomorphous replacement techniques. I. A distribution function for quartet invariants. Acta Cryst.* A**52**, 133–142.

Giacovazzo, C. & Siliqi, D. (1996b). *On integrating direct methods and isomorphous replacement techniques. II. The quartet invariant estimate. Acta Cryst.* A**52**, 143–151.

Giacovazzo, C. & Siliqi, D. (2002). *The method of joint probability distribution functions applied to SIR–MIR and to SIRAS–MIRAS cases. Acta Cryst.* A**58**, 590–597.

Giacovazzo, C. & Siliqi, D. (2004). *Phasing via SAD/MAD data: the method of the joint probability distribution functions. Acta Cryst.* D**60**, 73–82.

Giacovazzo, C., Siliqi, D. & García-Rodríguez, L. (2001). *On integrating direct methods and isomorphous-replacement techniques: triplet estimation and treatment of errors. Acta Cryst.* A**57**, 571–575.

Gillis, J. (1948). *Structure factor relations and phase determination. Acta Cryst.* **1**, 76–80.

Gilmore, C. J. (1984). *MITHRIL. An integrated direct-methods computer program. J. Appl. Cryst.* **17**, 42–46.

Goedkoop, J. A. (1950). *Remarks on the theory of phase limiting inequalities and equalities. Acta Cryst.* **3**, 374–378.

Gramlich, V. (1984). *The influence of rational dependence on the probability distribution of structure factors. Acta Cryst.* A**40**, 610–616.

Grant, D. F., Howells, R. G. & Rogers, D. (1957). *A method for the systematic application of sign relations. Acta Cryst.* **10**, 489–497.

Grosse-Kunstleve, R. W. & Brunger, A. T. (1999). *A highly automated heavy-atom search procedure for macromolecular structures. Acta Cryst.* D**55**, 1568–1577.

Hall, S. R., du Boulay, D. J. & Olthof-Hazekamp, R. (1999). *Xtal3.6 crystallographic software.* http://xtal.sourceforge.net/.

Harker, D. & Kasper, J. S. (1948). *Phases of Fourier coefficients directly from crystal diffraction data. Acta Cryst.* **1**, 70–75.

Hauptman, H. (1964). *The role of molecular structure in the direct determination of phase. Acta Cryst.* **17**, 1421–1433.

Hauptman, H. (1965). *The average value of $\exp\{2\pi i(\mathbf{h} \cdot \mathbf{r} + \mathbf{h}' \cdot \mathbf{r}')\}$. Z. Kristallogr.* **121**, 1–8.

Hauptman, H. (1970). Communication at New Orleans Meeting of Am. Crystallogr. Assoc.

Hauptman, H. (1974). *On the identity and estimation of those cosine invariants, $\cos(\varphi_\mathbf{m} + \varphi_\mathbf{n} + \varphi_\mathbf{p} + \varphi_\mathbf{q})$, which are probably negative. Acta Cryst.* A**30**, 472–476.

Hauptman, H. (1975). *A new method in the probabilistic theory of the structure invariants. Acta Cryst.* A**31**, 680–687.

Hauptman, H. (1982a). *On integrating the techniques of direct methods and isomorphous replacement. I. The theoretical basis. Acta Cryst.* A**38**, 289–294.

Hauptman, H. (1982b). *On integrating the techniques of direct methods with anomalous dispersion. I. The theoretical basis. Acta Cryst.* A**38**, 632–641.

Hauptman, H., Fisher, J., Hancock, H. & Norton, D. A. (1969). *Phase determination for the estriol structure. Acta Cryst.* B**25**, 811–814.

Hauptman, H. & Green, E. A. (1976). *Conditional probability distributions of the four-phase structure invariant $\varphi_\mathbf{h} + \varphi_\mathbf{k} + \varphi_\mathbf{l} + \varphi_\mathbf{m}$ in P$\bar{1}$. Acta Cryst.* A**32**, 45–49.

Hauptman, H. & Green, E. A. (1978). *Pairs in P2$_1$: probability distributions which lead to estimates of the two-phase structure seminvariants in the vicinity of $\pi/2$. Acta Cryst.* A**34**, 224–229.

Hauptman, H. & Karle, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* Am. Crystallogr. Assoc. Monograph No. 3. Dayton, Ohio: Polycrystal Book Service.

Hauptman, H. & Karle, J. (1956). *Structure invariants and seminvariants for non-centrosymmetric space groups. Acta Cryst.* **9**, 45–55.

Hauptman, H. & Karle, J. (1958). *Phase determination from new joint probability distributions: space group P$\bar{1}$. Acta Cryst.* **11**, 149–157.

Hauptman, H. & Karle, J. (1959). *Table 2. Equivalence classes, seminvariant vectors and seminvariant moduli for the centered centrosymmetric space groups, referred to a primitive unit cell. Acta Cryst.* **12**, 93–97.

Hauptman, H. A. (1991). In *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry. IUCr/Oxford University Press.

Hauptman, H. A., Xu, H., Weeks, C. M. & Miller, R. (1999). *Exponential Shake-and-Bake: theoretical basis and applications. Acta Cryst.* A**55**, 891–900.

Heinermann, J. J. L. (1977*a*). *The use of structural information in the phase probability of a triple product. Acta Cryst.* A**33**, 100–106.

Heinermann, J. J. L. (1977*b*). Thesis. University of Utrecht.

Heinermann, J. J. L., Krabbendam, H. & Kroon, J. (1979). *The joint probability distribution of the structure factors in a Karle–Hauptman matrix. Acta Cryst.* A**35**, 101–105.

Heinermann, J. J. L., Krabbendam, H., Kroon, J. & Spek, A. L. (1978). *Direct phase determination of triple products from Bijvoet inequalities. II. A probabilistic approach. Acta Cryst.* A**34**, 447–450.

Hendrickson, W. A. (1985). *Analysis of protein structure from diffraction measurement at multiple wavelengths. Trans. Am. Crystallogr. Assoc.* **21**, 11–21.

Hendrickson, W. A., Love, W. E. & Karle, J. (1973). *Crystal structure analysis of sea lamprey hemoglobin at 2 Å resolution. J. Mol. Biol.* **74**, 331–361.

Hoppe, W. (1963). *Phase determination and zero points in the Patterson function. Acta Cryst.* **16**, 1056–1057.

Hughes, E. W. (1953). *The signs of products of structure factors. Acta Cryst.* **6**, 871.

Hull, S. E. & Irwin, M. J. (1978). *On the application of phase relationships to complex structures. XIV. The additional use of statistical information in tangent-formula refinement. Acta Cryst.* A**34**, 863–870.

Hull, S. E., Viterbo, D., Woolfson, M. M. & Shao-Hui, Z. (1981). *On the application of phase relationships to complex structures. XIX. Magic-integer representation of a large set of phases: the MAGEX procedure. Acta Cryst.* A**37**, 566–572.

International Tables for Crystallography (2001). Vol. F, *Macromolecular Crystallography*, edited by M. G. Rossmann & E. Arnold. Dordrecht: Kluwer Academic Publishers.

Jaynes, E. T. (1957). *Information theory and statistical mechanics. Phys. Rev.* **106**, 620–630.

Karle, J. (1970*a*). *An alternative form for $B_{3.0}$, a phase determining formula. Acta Cryst.* B**26**, 1614–1617.

Karle, J. (1970*b*). *Partial structures and use of the tangent formula and translation functions.* In *Crystallographic Computing*, pp. 155–164. Copenhagen: Munksgaard.

Karle, J. (1972). *Translation functions and direct methods. Acta Cryst.* B**28**, 820–824.

Karle, J. (1979). *Triple phase invariants: formula for centric case from fourth-order determinantal joint probability distributions. Proc. Natl Acad. Sci. USA*, **76**, 2089–2093.

Karle, J. (1980*a*). *Triplet phase invariants: formula for acentric case from fourth-order determinantal joint probability distributions. Proc. Natl Acad. Sci. USA*, **77**, 5–9.

Karle, J. (1980*b*). *Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. Int. J. Quantum Chem. Quantum Biol. Symp.* **7**, 357–367.

Karle, J. (1983). *A simple rule for finding and distinguishing triplet phase invariants with values near 0 or π with isomorphous replacement data. Acta Cryst.* A**39**, 800–805.

Karle, J. & Hauptman, H. (1950). *The phases and magnitudes of the structure factors. Acta Cryst.* **3**, 181–187.

Karle, J. & Hauptman, H. (1956). *A theory of phase determination for the four types of non-centrosymmetric space groups $1P222$, $2P22$, $3P_12$, $3P_22$. Acta Cryst.* **9**, 635–651.

Karle, J. & Hauptman, H. (1958). *Phase determination from new joint probability distributions: space group $P1$. Acta Cryst.* **11**, 264–269.

Karle, J. & Hauptman, H. (1961). *Seminvariants for non-centrosymmetric space groups with conventional centered cells. Acta Cryst.* **14**, 217–223.

Karle, J. & Karle, I. L. (1966). *The symbolic addition procedure for phase determination for centrosymmetric and non-centrosymmetric crystals. Acta Cryst.* **21**, 849–859.

Klug, A. (1958). *Joint probability distributions of structure factors and the phase problem. Acta Cryst.* **11**, 515–543.

Koch, M. H. J. (1974). *On the application of phase relationships to complex structures. IV. Automatic interpretation of electron-density maps for organic structures. Acta Cryst.* A**30**, 67–70.

Krabbendam, H. & Kroon, J. (1971). *A relation between structure factor, triple products and a single Patterson vector, and its application to sign determination. Acta Cryst.* A**27**, 362–367.

Kroon, J., Spek, A. L. & Krabbendam, H. (1977). *Direct phase determination of triple products from Bijvoet inequalities. Acta Cryst.* A**33**, 382–385.

Kyriakidis, C. E., Peschar, R. & Schenk, H. (1996). *The estimation of four-phase structure invariants using the single difference of isomorphous structure factors. Acta Cryst.* A**52**, 77–87.

Lajzérowicz, J. & Lajzérowicz, J. (1966). *Loi de distribution des facteurs de structure pour un répartition non uniforme des atomes. Acta Cryst.* **21**, 8–12.

Langs, D. A. (1985). *Translation functions: the elimination of structure-dependent spurious maxima. Acta Cryst.* A**41**, 305–308.

Lessinger, L. & Wondratschek, H. (1975). *Seminvariants for space groups $I\bar{4}2m$ and $I\bar{4}d$. Acta Cryst.* A**31**, 521.

Mackay, A. L. (1953). *A statistical treatment of superlattice reflexions. Acta Cryst.* **6**, 214–215.

Main, P. (1976). *Recent developments in the MULTAN system. The use of molecular structure.* In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 97–105. Copenhagen: Munksgaard.

Main, P. (1977). *On the application of phase relationships to complex structures. XI. A theory of magic integers. Acta Cryst.* A**33**, 750–757.

Main, P., Fiske, S. J., Germain, G., Hull, S. E., Declercq, J.-P., Lessinger, L. & Woolfson, M. M. (1999). *Crystallographic software: teXsan for Windows.* http://www.rigaku.com/downloads/journal/Vol15.1.1998/texsan.pdf.

Main, P. & Hull, S. E. (1978). *The recognition of molecular fragments in E maps and electron density maps. Acta Cryst.* A**34**, 353–361.

Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *SnB: crystal structure determination via shake-and-bake. J. Appl. Cryst.* **27**, 613–621.

Mukherjee, A. K., Helliwell, J. R. & Main, P. (1989). *The use of MULTAN to locate the positions of anomalous scatterers. Acta Cryst.* A**45**, 715–718.

Narayan, R. & Nityananda, R. (1982). *The maximum determinant method and the maximum entropy method. Acta Cryst.* A**38**, 122–128.

Navaza, J. (1985). *On the maximum-entropy estimate of the electron density function. Acta Cryst.* A**41**, 232–244.

Navaza, J., Castellano, E. E. & Tsoucaris, G. (1983). *Constrained density modifications by variational techniques. Acta Cryst.* A**39**, 622–631.

Naya, S., Nitta, I. & Oda, T. (1964). *A study on the statistical method for determination of signs of structure factors. Acta Cryst.* **17**, 421–433.

Naya, S., Nitta, I. & Oda, T. (1965). *Affinement tridimensional du sulfanilamide β. Acta Cryst.* **19**, 734–747.

Nordman, C. E. (1985). *Introduction to Patterson search methods.* In *Crystallographic Computing 3. Data Collection, Structure Determination, Proteins and Databases*, edited by G. M. Sheldrick, G. Kruger & R. Goddard, pp. 232–244. Oxford: Clarendon Press.

Oda, T., Naya, S. & Taguchi, I. (1961). *Matrix theoretical derivation of inequalities. II. Acta Cryst.* **14**, 456–458.

Okaya, J. & Nitta, I. (1952). *Linear structure factor inequalities and the application to the structure determination of tetragonal ethylenediamine sulphate. Acta Cryst.* **5**, 564–570.

Okaya, Y. & Pepinsky, R. (1956). *New formulation and solution of the phase problem in X-ray analysis of non-centric crystals containing anomalous scatterers. Phys. Rev.* **103**, 1645–1647.

Okaya, Y., Saito, Y. & Pepinsky, R. (1955). *New method in X-ray crystal structure determination involving the use of anomalous dispersion. Phys. Rev.* **98**, 1857–1858.

Oszlányi, G. & Süto, A. (2004). *Ab initio structure solution by charge flipping. Acta Cryst.* A**60**, 134–141.

Ott, H. (1927). *Zur Methodik der Strukturanalyse. Z. Kristallogr.* **66**, 136–153.

Pähler, A., Smith, J. L. & Hendrickson, W. A. (1990). *A probability representation for phase information from multiwavelength anomalous dispersion. Acta Cryst.* A**46**, 537–540.

Parthasarathy, S. & Srinivasan, R. (1964). *The probability distribution of Bijvoet differences. Acta Cryst.* **17**, 1400–1407.

Pavelcík, F. (1988). *Patterson-oriented automatic structure determination: getting a good start. Acta Cryst.* A**44**, 724–729.

Pavelčík, F., Kuchta, L. & Sivý, J. (1992). *Patterson-oriented automatic structure determination. Utilizing Patterson peaks. Acta Cryst.* A**48**, 791–796.

Peerdeman, A. F. & Bijvoet, J. M. (1956). *The indexing of reflexions in investigations involving the use of the anomalous scattering effect. Acta Cryst.* **9**, 1012–1015.

Perutz, M. F. (1956). *Isomorphous replacement and phase determination in non-centrosymmetric space groups. Acta Cryst.* **9**, 867–873.

Phillips, D. C. (1966). *Advances in protein crystallography.* In *Advances in Structure Research by Diffraction Methods*, Vol. 2, edited by R. Brill & R. Mason, pp. 75–140. New York: John Wiley.

Piro, O. E. (1983). *Information theory and the phase problem in crystallography. Acta Cryst.* A**39**, 61–68.

Podjarny, A. D., Schevitz, R. W. & Sigler, P. B. (1981). *Phasing low-resolution macromolecular structure factors by matricial direct methods. Acta Cryst.* A**37**, 662–668.

Podjarny, A. D., Yonath, A. & Traub, W. (1976). *Application of multivariate distribution theory to phase extension for a crystalline protein. Acta Cryst.* A**32**, 281–292.

Rae, A. D. (1977). *The use of structure factors to find the origin of an oriented molecular fragment. Acta Cryst.* A**33**, 423–425.

Ramachandran, G. N. & Raman, S. (1956). *A new method for the structure analysis of non-centrosymmetric crystals. Curr. Sci. (India)*, **25**, 348.

Rango, C. de (1969). Thesis. Paris.

Rango, C. de, Mauguen, Y. & Tsoucaris, G. (1975). *Use of high-order probability laws in phase refinement and extension of protein structures. Acta Cryst.* A**31**, 227–233.

Rango, C. de, Mauguen, Y., Tsoucaris, G., Dodson, E. J., Dodson, G. G. & Taylor, D. J. (1985). *The extension and refinement of the 1.9 Å spacing isomorphous phases to 1.5 Å spacing in 2Zn insulin by determinantal methods. Acta Cryst.* A**41**, 3–17.

Rango, C. de, Tsoucaris, G. & Zelwer, C. (1974). *Phase determination from the Karle–Hauptman determinant. II. Connexion between inequalities and probabilities. Acta Cryst.* A**30**, 342–353.

Richardson, J. W. & Jacobson, R. A. (1987). *Patterson and Pattersons*, edited by J. P. Glusker, B. K. Patterson & M. Rossi, pp. 310–317. Oxford University Press.

Rogers, D., Stanley, E. & Wilson, A. J. C. (1955). *The probability distribution of intensities. VI. The influence of intensity errors on the statistical tests. Acta Cryst.* **8**, 383–393.

Rogers, D. & Wilson, A. J. C. (1953). *The probability distribution of X-ray intensities. V. A note on some hypersymmetric distributions. Acta Cryst.* **6**, 439–449.

Rossmann, M. G. (1960). *The accurate determination of the position and shape of heavy-atom replacement groups in proteins. Acta Cryst.* **13**, 221–226.

Rossmann, M. G. (1961). *The position of anomalous scatterers in protein crystals. Acta Cryst.* **14**, 383–388.

Rossmann, M. G., Blow, D. M., Harding, M. M. & Coller, E. (1964). *The relative positions of independent molecules within the same asymmetric unit. Acta Cryst.* **17**, 338–342.

Sayre, D. (1952). *The squaring method: a new method for phase determination. Acta Cryst.* **5**, 60–65.

Sayre, D. (1953). *Double Patterson function. Acta Cryst.* **6**, 430–431.

Sayre, D. (1972). *On least-squares refinement of the phases of crystallographic structure factors. Acta Cryst.* A**28**, 210–212.

Sayre, D. & Toupin, R. (1975). *Major increase in speed of least-squares phase refinement. Acta Cryst.* A**31**, S20.

Schenk, H. (1973a). *Direct structure determination in P1 and other non-centrosymmetric symmorphic space groups. Acta Cryst.* A**29**, 480–481.

Schenk, H. (1973b). *The use of phase relationships between quartets of reflexions. Acta Cryst.* A**29**, 77–82.

Schneider, T. R. & Sheldrick, G. M. (2002). *Substructure solution with SHELXD. Acta Cryst.* D**58**, 1772–1779.

Sheldrick, G. M. (1990). *Phase annealing in SHELX-90: direct methods for larger structures. Acta Cryst.* A**46**, 467–473.

Sheldrick, G. M. (1992). *Crystallographic Computing 5*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 145–157. Oxford University Press.

Sheldrick, G. M. (1998). *SHELX: applications to macromolecules.* In *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 401–411. Dordrecht: Kluwer Academic Publishers.

Sheldrick, G. M. (2000). *The SHELX home page.* http://shelx.uni-ac.gwdg.de/SHELX/.

Sheldrick, G. M. & Gould, R. O. (1995). *Structure solution by iterative peaklist optimization and tangent expansion in space group P1. Acta Cryst.* B**51**, 423–431.

Sim, G. A. (1959). *The distribution of phase angles for structures containing heavy atoms. II. A modification of the normal heavy-atoms method for non-centrosymmetrical structures. Acta Cryst.* **12**, 813–815.

Simerska, M. (1956). *Czech. J. Phys.* **6**, 1.

Simonov, V. I. & Weissberg, A. M. (1970). *Calculation of the signs of structure amplitudes by a binary function section of interatomic vectors. Sov. Phys. Dokl.* **15**, 321–323.

Sint, L. & Schenk, H. (1975). *Phase extension and refinement in non-centrosymmetric structures containing large molecules. Acta Cryst.* A**31**, S22.

Srinivasan, R. & Parthasarathy, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon Press.

Taylor, D. J., Woolfson, M. M. & Main, P. (1978). *On the application of phase relationships to complex structures. XV. Magic determinants. Acta Cryst.* A**34**, 870–883.

Terwilliger, T. C. (1994). *MAD phasing: Bayesian estimates of F_A. Acta Cryst.* D**50**, 11–16.

Terwilliger, T. C. & Eisenberg, D. (1987). *Isomorphous replecement: effect of errors on the phase probability distribution. Acta Cryst.* A**43**, 6–13.

Tsoucaris, G. (1970). *A new method for phase determination. The maximum determinant rule. Acta Cryst.* A**26**, 492–499.

Van der Putten, N. & Schenk, H. (1977). *On the conditional probability of quintets. Acta Cryst.* A**33**, 856–858.

Vaughan, P. A. (1958). *A phase-determining procedure related to the vector-coincidence method. Acta Cryst.* **11**, 111–115.

Vermin, W. J. & de Graaff, R. A. G. (1978). *The use of Karle–Hauptman determinants in small-structure determinations. Acta Cryst.* A**34**, 892–894.

Vicković, I. & Viterbo, D. (1979). *A simple statistical treatment of unobserved reflexions. Application to two organic substances. Acta Cryst.* A**35**, 500–501.

Weeks, C. M., DeTitta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Structure solution by minimal-function phase refinement and Fourier filtering. II. Implementation and applications. Acta Cryst.* A**50**, 210–220.

Weeks, C. M. & Miller, R. (1999). *The design and implementation of SnB version 2.0. J. Appl. Cryst.* **32**, 120–124.

Weinzierl, J. E., Eisenberg, D. & Dickerson, R. E. (1969). *Refinement of protein phases with the Karle–Hauptman tangent fomula. Acta Cryst.* B**25**, 380–387.

White, P. & Woolfson, M. M. (1975). *The application of phase relationships to complex structures. VII. Magic integers. Acta Cryst.* A**31**, 53–56.

Wilkins, S. W., Varghese, J. N. & Lehmann, M. S. (1983). *Statistical geometry. I. A self-consistent approach to the crystallographic inversion problem based on information theory. Acta Cryst.* A**39**, 47–60.

Wilson, A. J. C. (1942). *Determination of absolute from relative X-ray intensity data. Nature (London)*, **150**, 151–152.

Wilson, K. S. (1978). *The application of MULTAN to the analysis of isomorphous derivatives in protein crystallography. Acta Cryst.* B**34**, 1599–1608.

Wolff, P. M. de & Bouman, J. (1954). *A fundamental set of structure factor inequalities. Acta Cryst.* **7**, 328–333.

Woolfson, M. M. (1958). *Crystal and molecular structure of p,p′-dimethoxybenzophenone by the direct probability method. Acta Cryst.* **11**, 277–283.

Woolfson, M. M. (1977). *On the application of phase relationships to complex structures. X. MAGLIN – a successor to MULTAN. Acta Cryst.* A**33**, 219–225.

Yao, J.-X. (1981). *On the application of phase relationships to complex structures. XVIII. RANTAN – random MULTAN. Acta Cryst.* A**37**, 642–664.

references